

**NOVEL 14275, 54420, 8797, 27439, 68730, 69112 AND 52908 MOLECULES AND
USES THEREFOR**

Related Applications:

[0001] The present application is a continuation-in-part of U.S. Patent Application Serial No. 10/007,399, filed November 5, 2001 (pending), which is a continuation of U.S. Patent Application Serial No. 09/390,039, filed September 3, 1999, which is a continuation-in-part of U.S. Patent Application Serial No. 09/146,416, filed September 3, 1998. The present application is also a continuation-in-part of U.S. Patent Application Serial No. 10/103,458, filed March 22, 2002 (pending), which is a continuation of U.S. Patent Application Serial No. 09/544,797, filed April 7, 2000. The present application is also a continuation-in-part of U.S. Patent Application Serial No. 09/945,254, filed August 31, 2001 (pending), which claims the benefit of U.S. Provisional Application Serial No. 60/229,829, filed on August 31, 2000. The present application is also a continuation-in-part of U.S. Patent Application Serial No. 09/945,301, filed August 31, 2001 (pending), which claims the benefit of U.S. Provisional Application Serial No. 60/229,301, filed September 1, 2000. The present application is also a continuation-in-part of U.S. Patent Application Serial No. 10/024,036, filed December 17, 2001 (pending), which claims the benefit of U.S. Provisional Application Serial No. 60/258,222, filed December 22, 2000. The present application is also a continuation-in-part of U.S. Patent Application Serial No. 10/192,440, filed July 10, 2002 (pending), which claims the benefit of U.S. Provisional Application Serial No. 60/341,953, filed December 19, 2001, and U.S. Provisional Application Serial No. 60/304,243, filed July 10, 2001. The entire contents of each of the above-referenced patent applications are incorporated herein by this reference.

INDEX

Chapter	Page	Title
I.	2	14275 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR RELATED TO THE EDG RECEPTOR FAMILY
II.	83	54420, A NOVEL HUMAN CALCIUM CHANNEL
III.	166	8797, A NOVEL HUMAN GALACTOSYLTRANSFERASE AND USES THEREOF
IV.	260	27439, A NOVEL HUMAN HYDROXYLASE AND USES THEREFOR
V.	358	68730 AND 69112, PROTEIN KINASE MOLECULES AND USES THEREFOR
VI.	445	52908, A HUMAN POTASSIUM CHANNEL, AND USES THEREOF

I. 14275 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR RELATED TO THE EDG RECEPTOR FAMILY

FIELD OF THE INVENTION

[0002] The present invention relates to a newly identified member of the superfamily of G-protein-coupled receptors, and a new member of the EDG receptor family. The invention also relates to polynucleotides encoding the receptor. The invention further relates to methods using receptor polypeptides and polynucleotides as a target for diagnosis and treatment in receptor-mediated disorders. The invention further relates to drug-screening methods using the receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

[0003] G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have seven transmembrane segments. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

[0004] GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosa (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424(1992)), nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

[0005] The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the beta2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science* 258:597-603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

[0006] There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; however, they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci.*

USA 90:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

[0007] G proteins represent a family of heterotrimeric proteins composed of α , β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane domains. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the $\beta\gamma$ -subunits. The GTP-bound form of the α -subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenylyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α -subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference.

Lipid Ligands for GPCRs

[0008] Lysophospholipids have been shown to act on distinct G-protein-coupled receptors to serve a variety of overlapping biological functions. Lysophosphatidic acid (LPA) is an extracellular phospholipid that produces multiple cellular responses including cellular proliferation, inhibition of differentiation, cell surface fibronectin binding, tumor cell invasion, chemotaxis, Cl^- mediated membrane depolarization, increased tight junction permeability, myoblast differentiation, stimulation of fibroblast chemotaxis, acute loss of gap junctional communication, platelet aggregation, smooth muscle contraction, neurotransmitter release, stress fiber formation, cell rounding, and neurite retraction, among others. See, Moolenaar, W.H. *et al.*, *Curr. Opin. Cell Biol.* 9:168-173 (1997). LPA acts through G-protein-coupled receptors to evoke the multiple cellular responses. It is generated from activated platelets and can also be generated from microvesicles shed from blood cells challenged with inflammatory stimuli. It is one of the major mitogens found in blood serum.

[0009] The N1E-115 neuronal cell line shows morphological responses to LPA. LPA induces retraction of developing neurites and rounding of the cell body, changes driven

by contraction of the actomyosin system, regulated by the GTP binding protein Rho. See, Postma, *EMBO J.* 15:2388-2395 (1996).

[0010] In *Xenopus* oocytes, LPA elicits oscillatory Cl^- currents. Expression depends upon a high affinity LPA receptor having features common to members of the rhodopsin seven transmembrane receptor superfamily. An antisense oligonucleotide derived from the first 5-11 amino acids selectively inhibited expression of this receptor. See, Guo *et al.*, *Proc. Natl. Acad. Sci. USA* 93:14367-14372 (1996).

[0011] The intracellular biochemical signaling events that mediate the effects of LPA include stimulation of phospholipase C and consequent increases in cytoplasmic calcium concentration, inhibition of adenylyl cyclase, and activation of phosphatidylinositol-3-kinase, the Ras-Raf-MAP kinase cascade and Rho GTPase and Rho-dependent kinases. The Ras-Raf-MAP kinase and Rho pathways stimulate the transcription factors ternary complex factor and serum response factor, respectively. Ternary complex factors and serum response factors synergistically activate transcription of growth-related immediate early genes such as *c-fos* by binding to serum response element (SRE) in the promoters (Hill *et al.*, *Cell* 81:1159-1170 (1995)).

[0012] LPA receptors in fibroblasts couple to at least three distinct G-proteins: G_q , G_i , and G_{12-13} . Activation of G_q stimulates phospholipase C and consequent mobilization of intracellular calcium. Activation of G_i inhibits adenylyl cyclase and stimulates the Ras-Raf-MAP kinase pathway leading to transcriptional activation mediated by ternary complex factors. Activation of G_{12-13} stimulates Rho which leads to actin-based cytoskeleton changes and transcriptional activation mediated by serum response factor. The G_i and Rho-activated pathways synergistically stimulate transcription of many growth-related genes containing serum response elements in their promoters (An, *et al.*, *J. Biol. Chem.* 273:7906-7910 (1998)).

[0013] It has been reported that serum albumin contains about a dozen as yet unidentified lipids (methanol soluble) with LPA-like biological activity. See Postma, cited above.

[0014] Sphingolipids have also been reported to be involved in cell signaling. Ceramide (N-acyl-sphingosine), sphingosine and sphingosine-1-phosphate (S1P) are second messengers involved in various biological functions. Ceramide is involved in apoptosis. S1P is a platelet-derived lysosphingolipid that acts on cognate G-protein-coupled receptors to evoke multiple cellular responses. See Moolenaar, cited above, and Meyer *et al.*, *FEBS*.

Lett. 410:34-38 (1997) for a review. Typical receptor-mediated responses to S1P (and LPA) include stimulation of phospholipase C and consequent calcium mobilization, inhibition of adenylate cyclase, mitogen activated protein (MAP) kinase activation, DNA synthesis, mitogenesis and cytoskeletal changes, such as cell rounding and neurite retraction (Zondag, cited above), microfilament reorganization, cell migration, stress fiber formation, membrane depolarization, and fibroblast proliferation.

[0015] S1P has been shown to act on neuronal N1E-115 cells by means of a high affinity receptor, to remodel the actin cytoskeleton in a Rho-dependent manner. See, Postma, *et al.*, cited above. Like LPA, S1P induces neurite retraction and cell rounding in differentiated PC12 cells. See, Sato *et al.*, *Biochem. Biophys. Res. Comm.* 240:329-334 (1997).

[0016] S1P acts by activating a G-protein-coupled receptor distinct from the LPA receptor.

[0017] A distinct receptor is also activated by another lysosphingolipid, sphingosyl-phosphorylcholine (SPC or lysosphingomyelin). It is a strong mitogen and evokes biochemical responses similar to those by LPA, except by a distinct receptor (in some cells, however, SPC and S1P might act on the same receptor). See, Moolenaar, cited above. SPC has also been shown to mediate fibroblast mitogenesis, platelet activation, and neurite retraction. It has been shown to activate MAP kinases. See, An *et al.*, *FEBS Lett.* 417:279-282 (1997). S1P and SPC also activate pathways involving G_i, Ras-Raf-ERK and Rho GTPases (An *et al.*, *FEBS Lett.*).

[0018] Since S1P and LPA are both released from activated platelets, they may play a role in wound healing and tissue remodeling, including during traumatic injury of the nervous system. Because LPA can also be generated from blood cells challenged with inflammatory stimuli, LPA may stimulate responses not only at the site of injury but also at sites of inflammation.

EDG receptors

[0019] Hecht *et al.* (*J. Cell Biol.* 135:1071-1083 (1996)) cloned a cDNA from mouse neocortical cell lines. This gene, termed ventricular zone gene-1 (*vzg-1*) was shown to be 96% identical to an unpublished sheep sequence designated EDG-2 (GenBank Accession No. U18405) and identified as an LPA receptor. This cDNA was also isolated as an orphan receptor by Macrae *et al.* (*Mol. Brain Res.* 42:245-254 (1996)) who designated it Rec1.3.

EDG-2 is closely homologous to a G_i-linked orphan receptor EDG-1 (37% homology). A cDNA homologous to that encoding sheep EDG-2 protein was cloned from a human lung cDNA library (An *et al.*, *Biochem. Biophys. Res. Comm.* 231:619-622 (1997)). A search of GenBank showed that EDG-2 cDNA from mouse and cow had also been cloned and sequenced. The human EDG-2 protein was shown to be a receptor for LPA. The cDNA was expressed in mammalian cells (HEK293 and CHO) using a reporter gene assay quantifying the transcriptional activation of a serum response element-containing promoter. This assay can sensitively measure the G-protein-activated signaling pathways linked to LPA receptors. The mouse EDG-2 (*Vzg-1*) showed 96% identity to the human EDG-2 (Hecht *et al.*, *J. Cell Biol.* 135:1071-1083 (1996)). EDG-2 was demonstrated to mediate inhibition of adenyl cyclase by Gi and cell morphological changes via Rho-related GTPases (An *et al.*, *J. Biol. Chem.* 273:7906-7910 (1998)).

[0020] Human EDG-1 cDNA was cloned from a human cDNA library of human umbilical vein endothelial cells exposed to fluid sheer stress (Takada *et al.*, *Biochem. Biophys. Res. Comm.* 240:737-741 (1997)). EDG-1 mRNA levels in endothelial cells increased markedly in response to fluid flow. This suggested that EDG-1 is a receptor gene that could function to regulate endothelial function under physiological blood flow conditions. Recently, it was shown that the EDG-1 receptor is capable of mediating a subset of early responses to sphingosine 1-phosphate (S1P), notably, inhibition of adenylate cyclase and activation of the G_i-MAP kinase pathway, but not activation of the PLC-Ca²⁺ signaling pathway. (Zondag, G.C. *et al.*, *Bio. Chem. J.* 330:605-609 (1998)).

[0021] In the study of Zondag, the results indicated that EDG-1 but not EDG-2 was capable of mediating the specific subset of cellular actions induced by S1P. However, these responses were specific in that LPA failed to mimic S1P.

[0022] Another study (Fukushima *et al.*, *Proc. Nat'l. Acad. Sci. USA* 95:6151-6156 (1998)) showed that the human EDG-2 mediates multiple cellular responses to LPA. At least six biological responses to LPA were reported, including the production of LPA membrane binding sites, LPA dependent G-protein activation, stress fiber formation, neurite retraction, transcriptional serum response element activation and increased DNA synthesis. EDG-1 and EDG-2 were shown to signal through at least two distinct pathways, a G_i/G_o pathway and a PTX insensitive pathway that involves Rho activation. It was demonstrated that G_i coupled directly with *Vzg-1* (EDG-2) after LPA exposure. At the same time it was shown that *Vzg-1* mediates actin-based cytoskeletal changes that operate through a Rho-

sensitive pathway. See Fukushima, cited above. The results were consistent with a model in which EDG-2 transduces LPA signals onto the same DNA target through two separate pathways. Activation of serum response element-dependent transcription can be effected through stimulation of the Ras-Raf-MAP kinase cascade (by a ternary complex factor) and through a Rho-mediated pathway. An important response related to the serum response element activation is progression through the cell cycle.

[0023] Using the cDNA sequence of the EDG-2 human LPA receptor to perform a sequence-based search for lysosphingolipid receptors, An *et al.* (*FEBS. Lett.* 417:279-282 (1997)) found two closely related G-protein-coupled receptors, designated rat H218 and human EDG-3. Both of these, when overexpressed in Jurkat cells, mobilized calcium and activated serum response element-driven transcriptional reporter gene (which requires activation of Rho and Ras GTPases) in response to S1P, dihydro-S1P, and sphingosylphosphorylcholine, but not to LPA. Expressed in *Xenopus* oocytes, the genes conferred responsiveness to S1P in agonist-triggered calcium efflux.

[0024] EDG-2 was also used for a sequence-based search for new genes encoding novel subtypes of LPA receptors. A human cDNA encoding a G-protein-coupled receptor designated EDG-4 was identified by searching GenBank for homologies with the EDG-2 LPA receptor. When overexpressed in Jurkat cells, this protein mediates LPA-induced activation of a serum response element reporter gene with LPA concentration-dependence and specificity (An *et al.*, *J. Biol. Chem.* 273:7906-7910 (1998)). Jurkat cells are a preferred assay system because they lack background responses to LPA in the serum response element reporter gene assay. EDG4 was shown to mediate activation of serum response element-driven transcription in Jurkat cells involving G_i and Rho GTPase.

[0025] GPCRs in general and EDG receptors are important targets for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs, particularly EDG receptors. The present invention advances the state of the art by providing a previously unidentified human GPCR, a new member of the EDG receptor family.

SUMMARY OF THE INVENTION

[0026] It is an object of the invention to identify novel GPCRs.

[0027] It is a further object of the invention to provide novel GPCR polypeptides that are useful as reagents or targets in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders.

[0028] It is a further object of the invention to provide polynucleotides corresponding to the novel GPCR polypeptides that are useful as targets and reagents in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders and useful for producing novel receptor polypeptides by recombinant methods.

[0029] A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the receptor.

[0030] A further specific object of the invention is to provide the compounds that modulate the expression of the receptor for treatment and diagnosis of GPCR related disorders.

[0031] The invention is thus based on the identification of a novel GPCR, designated the 14275 receptor.

[0032] The invention provides isolated 14275 receptor polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO:2.

[0033] The invention also provides isolated 14275 receptor nucleic acid molecules having the sequence shown in SEQ ID NO:1.

[0034] The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO:2.

[0035] The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO:1.

[0036] The invention also provides fragments of the polypeptide shown in SEQ ID NO:2 and nucleotide shown in SEQ ID NO:1, as well as substantially homologous fragments of the polypeptide or nucleic acid.

[0037] The invention further provides nucleic acid constructs comprising the nucleic acid molecules described herein. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

[0038] The invention also provides vectors and host cells for expression of the receptor nucleic acid molecules and polypeptides and particularly recombinant vectors and host cells.

[0039] The invention also provides methods of making the vectors and host cells and methods for using them to produce the receptor nucleic acid molecules and polypeptides.

[0040] The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the receptor polypeptides and fragments.

[0041] The invention also provides methods of screening for compounds that modulate expression or activity of the polypeptides or nucleic acid (RNA or DNA).

[0042] The invention also provides a process for modulating polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the polypeptides or nucleic acids.

[0043] The invention also provides assays for determining the presence or absence of and level of the polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[0044] The invention also provides assays for determining the presence of a mutation in the polypeptides or nucleic acid molecules, including for disease diagnosis.

[0045] In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention, respectively.

[0046] The invention also provides methods of screening for compounds that modulate the activity of the receptor polypeptides. Modulation can be at the level of the polypeptide receptor or at the level of controlling the expression of nucleic acid expressing the receptor polypeptide.

[0047] The invention also provides a process for modulating receptor polypeptide activity, especially using the screened compounds, including to treat conditions related to expression of the receptor polypeptides.

[0048] The invention also provides diagnostic assays for determining the presence of and level of the receptor polypeptides or nucleic acid molecules in a biological sample.

[0049] The invention also provides diagnostic assays for determining the presence of a mutation in the receptor polypeptides or nucleic acid molecules.

DETAILED DESCRIPTION OF THE INVENTION

Receptor function/signal pathway

[0050] The 14275 receptor protein is a GPCR that participates in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (14275 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP₂), inositol 1,4,5-triphosphate (IP₃) or adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. Functions mediated by EDG receptors are further presented in the background section, *supra*.

[0051] Since the 14275 receptor protein is expressed in peripheral blood cells, spleen, lung, small intestine, prostate, heart, thymus, colon, uterus and placenta, cells participating in a 14275 receptor protein signaling pathway include, but are not limited to cells derived from these tissues.

[0052] Depending on the type of cell, the response mediated by the receptor protein may be different. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that the protein is a GPCR and interact with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

[0053] As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP₂) as well as to the activities of these molecules. PIP₂ is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP₂ to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP₃). Once formed IP₃ can diffuse to the endoplasmic reticulum surface where it can bind an IP₃ receptor, e.g., a

calcium channel protein containing an IP₃ binding site. IP₃ binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP₃ can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP₄), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP₃ and IP₄ can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP₂) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP₂. The other second messenger produced by the hydrolysis of PIP₂, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF- κ B. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP₂ or one of its metabolites.

[0054] Another signaling pathway the receptor may participate in is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

Polypeptides

[0055] The invention is based on the identification of a novel G-coupled protein receptor. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences. This EST was used to design primers based on sequences that it contains and used to identify a cDNA from a human prostate cDNA

library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequence revealed that the cloned cDNA molecule encodes a G-protein coupled receptor showing a high homology score against the seven transmembrane segment rhodopsin superfamily, also with high homology to the EDG receptor family.

[0056] The invention thus relates to a novel GPCR having the deduced amino acid sequence of SEQ ID NO:2.

[0057] The "14275 receptor polypeptide" or "14275 receptor protein" refers to the polypeptide in SEQ ID NO:2. The term "receptor protein" or "receptor polypeptide", however, further includes the numerous variants described herein, as well as fragments derived from the full length 14275 polypeptide and variants.

[0058] The present invention thus provides an isolated or purified 14275 receptor polypeptide and variants and fragments thereof.

[0059] The 14275 polypeptide is a 384 residue protein exhibiting three main structural domains. The amino terminal extracellular domain is identified to be within residues 1 to about 50 in SEQ ID NO:2. The region spanning the entire transmembrane domain is identified by hydrophobicity plot to be within residues from about 51 to about 331 in SEQ ID NO:2. The transmembrane domain includes the seven transmembrane segments, the three intracellular loops and the three extracellular loops. Discrete transmembrane segments are estimated to be from about amino acid 51-71, 81-105, 123-141, 162-184, 204-227, 253-276, and 291-331. Accordingly, the six extracellular and intracellular loops correspond to about amino acids 106-122, 185-203, 277-290, 72-80, 142-161, and 228-252. The carboxy terminal intracellular domain is identified to be within residues from about 332 to about 384 in SEQ ID NO:2.

[0060] A glycosylation site is found from amino acids 2 to 5, which is in the amino terminal extracellular domain. A second glycosylation site is found at amino acids 30 to 33, which is also in the amino terminal extracellular domain. A third glycosylation site is found at amino acids 87 to 90, which is in the second transmembrane segment. A protein kinase C phosphorylation site is found at amino acids 32 to 34, which is in the amino terminal extracellular domain. A second protein kinase C phosphorylation site is found at amino acids 77 to 79, which is in the first intracellular loop. A third protein kinase C phosphorylation site is found at amino acids 110 to 112, which is in the first extracellular loop. A fourth protein kinase C phosphorylation site is found at amino acids 159 to 161, which is in the second intracellular loop. A fifth protein kinase C phosphorylation site is

found at amino acids 201 to 203, which is in the second extracellular loop. A sixth protein kinase C phosphorylation site is found at amino acids 308 to 310, which is in the seventh transmembrane segment. A seventh protein kinase C phosphorylation site is found at amino acids 354 to 356, which is in the carboxy terminal intracellular domain. An eighth protein kinase C phosphorylation site is found at amino acids 360 to 362, which is in the carboxy terminal intracellular domain. A ninth protein kinase C phosphorylation site is found at amino acids 368 to 370, which is in the carboxy terminal intracellular domain. A tenth protein kinase C phosphorylation site is found at amino acids 380 to 382, which is in the carboxy terminal intracellular domain. A casein kinase II phosphorylation site is found at amino acids 89 to 92, which is in the second transmembrane segment. A second casein kinase II phosphorylation site is found at amino acids 139 to 142, which spans the third transmembrane segment and second intracellular loop. A third casein kinase II phosphorylation site is found at amino acids 349 to 352, which is in the carboxy terminal intracellular domain. An N-myristoylation site is found at amino acids 44 to 49, which is in the amino terminal extracellular domain. A second N-myristoylation site is found at amino acids 51 to 56, which is in the first transmembrane segment. A third N-myristoylation site is found at amino acids 123 to 128, which is in the third transmembrane segment. A fourth N-myristoylation site is found at amino acids 155 to 160, which is in the second intracellular loop. A fifth N-myristoylation site is found at amino acids 214 to 219, which is in the fifth transmembrane segment. A sixth N-myristoylation site is found at amino acids 221 to 226, which is in the fifth transmembrane segment. A seventh N-myristoylation site is found at amino acids 269 to 274, which is in the sixth transmembrane segment. An eighth N-myristoylation site is found at amino acids 347 to 352, which is in the carboxy terminal intracellular domain. In addition, amino acids corresponding in position to the GPCR signature and containing the invariant arginine are found in the sequence ERF at amino acids 142 to 144.

[0061] A search which was performed against the HMM database resulted in the identification of a “7 transmembrane receptor (rhodopsin family)” domain (SEQ ID NO:3) at amino acids 63 to 307 of SEQ ID NO:2. The most commonly conserved intracellular sequence in the family is the aspartate, arginine, tyrosine (DRY) triplet. Arginine is invariant. Aspartate is conservatively placed in several GPCRs. DRY is implicated in signal transduction. In the present case, the arginine is found in the sequence ERF, which matches the position of the DRY or invariant arginine for a rhodopsin family seven

transmembrane receptor. The transmembrane domain includes the invariant arginine of a GPCR signal transduction signature, ERF, at amino acids 142 to 144.

[0062] One of skill in the art can use the Peptide Structure program in the GCG software package to find α turn and coil regions, hydrophilicity, amphipathic regions, flexible regions, antigenic index, and surface probability.

[0063] The 14275 amino acid sequence showed approximately 40% identity with EDG-4, 37% identity with EDG-2, 42% identity with EDG-3, and 43% identity with EDG-1. The approximate percent identity among various EDG family members are as follows: EDG1-EDG2:40%; EDG1-EDG4:40%; EDG1-EDG3:55%; EDG1-14275 receptor:43%; EDG2-EDG4:57%; EDG2-EDG3:39%; EDG2-14275 receptor:37%; EDG3-EDG4:32%; EDG3-14275 receptor:42%; EDG4-14275 receptor:40%.

[0064] As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

[0065] The receptor polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

[0066] In one embodiment, the language "substantially free of cellular material" includes preparations of the receptor polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the receptor polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

[0067] A polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

[0068] The language "substantially free of chemical precursors or other chemicals" includes preparations of the receptor polypeptide in which it is separated from chemical

precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

[0069] In one embodiment, the receptor polypeptide comprises the amino acid sequence shown in SEQ ID NO:2. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. In the present case, the 14725 receptor gene has been mapped to chromosome 2, near the marker NRB733. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 14275 receptor protein of SEQ ID NO:2. Variants also include proteins substantially homologous to the 14275 receptor protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 14275 receptor protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 14275 receptor protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

[0070] As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 55-60%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO:1 under stringent conditions as more fully described below.

[0071] To determine the percent homology of two amino acid sequences, or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence, then the molecules are homologous at that position. As used

herein, amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity". The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., per cent homology equals the number of identical positions/total number of positions times 100).

[0072] The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the 14275 polypeptide. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine

Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

[0073] Both identity and similarity can be readily calculated (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Preferred computer program methods to determine identify and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J. Molec. Biol.* 215:403 (1990)).

[0074] To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences ($\%, \text{ \% identity} = \frac{\text{\# of identical positions}}{\text{total \# of positions}} \times 100$) (*e.g.*, overlapping positions). In one embodiment, the two sequences are the same length.

[0075] The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of

Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used.

[0076] Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the CGC sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti (1994) *Comput. Appl. Biosci.*, 10:3-5; and FASTA described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-8. Within FASTA, ktup is a control option that sets the sensitivity and speed of the search. If ktup=2, similar regions in the two sequences being compared are found by looking at pairs of aligned residues; if ktup=1, single aligned amino acids are examined. ktup can be set to 2 or 1 for protein sequences, or from 1 to 6 for DNA sequences. The default if ktup is not specified is 2 for proteins and 6 for DNA. For a further description of FASTA parameters, see bioweb.pasteur.fr/docs/man/man/fasta.1.html#sect2, the contents of which are incorporated herein by reference.

[0077] The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

[0078] A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

[0079] Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to ligand binding, membrane association, G-protein binding and signal transduction.

[0080] Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

[0081] Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

[0082] As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the receptor polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

[0083] Useful variations further include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation includes one that allows binding but which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or sub-regions is operationally fused to one or more domains or sub-regions from another G-protein coupled receptor.

[0084] Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)). The latter procedure introduces single

alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al. Science* 255:306-312 (1992)).

[0085] The invention also includes polypeptide fragments of the 14275 receptor protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO:2. However, the invention also encompasses fragments of the variants of the 14275 receptor protein as described herein.

[0086] The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

[0087] Fragments can retain one or more of the biological activities of the protein, for example the ability to bind to a G-protein or ligand. Fragments can also be useful as an immunogen to generate receptor antibodies.

[0088] Biologically active fragments can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, or GPCR signature, glycosylation sites, protein kinase C, or casein kinase II phosphorylation sites, and myristoylation sites. Such peptides can be, for example, 7, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length. Such domains or motifs can be identified by means of routine computerized homology searching procedures.

[0089] Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain about amino acid 1 to about amino acid 50 of SEQ ID NO:2, or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain from about amino acid 332 to amino acid 384 of SEQ ID NO:2, or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain from about amino acid 51 to about amino acid 331, or parts thereof; 4) any of the specific transmembrane segments, or parts thereof, from about amino acid 51 to about amino acid 71, from about amino acid 81 to about amino acid 105, from about amino acid 123 to about amino acid 141, from about amino acid 162 to about amino acid 184, from about amino acid 204 to about amino acid 227, from about amino acid 253 to about amino acid 276, and from about amino acid 291 to about amino acid 331; 5) any of the three

intracellular or three extracellular loops, or parts thereof, from about amino acid 72 to about amino acid 80, from about amino acid 142 to about amino acid 161, from about amino acid 228 to about amino acid 252, from about amino acid 106 to about amino acid 122, from about amino acid 185 to about amino acid 203, and from about amino acid 277 to about amino acid 290. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to 384. Other fragments contain the various functional sites described herein, such as phosphorylation sites, glycosylation sites, and myristoylation sites and a sequence containing the GPCR signature sequence. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived. Fragments also include amino acid sequences greater than 17 amino acids. Fragments also include antigenic fragments and specifically those shown to have a high antigenic index. Further specific fragments include a fragment from about 1 to 120 and sub-fragments thereof greater than 6 amino acids, from about 116 to 296 and sub-fragments thereof greater than 9 amino acids, from about 288 to 361 and sub-fragments thereof greater than 10 amino acids, from about 352-384 and subfragments thereof greater than 17 amino acids, and from about 375 to 384 and sub-fragments thereof. Further fragments include a fragment including any amino acid sequences from 1-120 but extending beyond amino acid 120.

[0090] Accordingly, possible fragments include fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining phosphorylation sites, fragments defining interaction with G proteins and signal transduction, and fragments defining myristoylation sites. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site. These regions can be identified by well-known methods involving computerized homology analysis.

[0091] The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the 14275 receptor protein and variants. These epitope-bearing peptides are useful to raise antibodies that bind specifically to a receptor polypeptide or region or fragment. These peptides can contain at least 6, 9, 12, 14, or between at least about 15 to about 30 amino acids.

[0092] Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include peptides derived from the amino terminal extracellular domain or any of the extracellular loops. However, other peptides are possible, for example, intracellular regions that could serve as an intrabody target. Regions having a high antigenicity index can be found using the Peptide Structure program in the GCG software package.

[0093] The receptor polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR related conditions.

[0094] The epitope-bearing receptor and polypeptides may be produced by any conventional means (Houghten, R.A., *Proc. Nat'l. Acad. Sci. USA* 82:5131-5135 (1985)). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

[0095] Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the receptor fragment and an additional region fused to the carboxyl terminus of the fragment.

[0096] The invention thus provides chimeric or fusion proteins. These comprise a receptor protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the receptor protein. "Operatively linked" indicates that the receptor protein and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the receptor protein.

[0097] In one embodiment the fusion protein does not affect receptor function per se. For example, the fusion protein can be a GST-fusion protein in which the receptor sequences are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant receptor protein.

In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus.

[0098] EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.* (*Journal of Molecular Recognition* 8:52-58 (1995)) and Johanson *et al.* (*Journal of Biological Chemistry* 270 (16):9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a receptor polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated and can be cleaved with factor Xa.

[0100] A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A receptor protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the receptor protein.

[0101] Another form of fusion protein is one that directly affects receptor functions. Accordingly, a receptor polypeptide encompassed by the present invention in which one or more of the receptor domains (or parts thereof) has been replaced by homologous domains

(or parts thereof) from another G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) may be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the region spanning the entire transmembrane domain or any of the seven segments or loops, for example, G-protein-binding/signal transduction, may be replaced. Finally, the carboxy terminal intracellular domain or sub-region may be replaced. Thus, chimeric receptors can be formed in which one or more of the native domains or subregions has been replaced.

[0102] The isolated receptor protein can be purified from cells that naturally express it, such as from peripheral blood cells, such as T or B cells, mobilized peripheral blood CD34⁺ cells, in HL60 (promyelocytic leukemia) cell line, CD34⁺ mobilized bone marrow cells, CD8⁺T lymphocytes, CD34⁺ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, CD34⁺ mobilized peripheral blood cells, CD4⁺T lymphocytes, spleen, lung, thymus, uterus, small intestine, colon, heart, prostate, and placenta, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

[0103] In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the receptor polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

[0104] Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

[0105] Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature

polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

[0106] Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

[0107] Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663:48-62 (1992).

[0108] As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

[0109] Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

[0110] The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

[0111] The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide uses

[0112] The receptor polypeptides are useful for producing antibodies specific for the 14275 receptor protein, regions, or fragments. Regions having a high antigenicity index can be found.

[0113] The receptor polypeptides (including variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions, especially disorders involving the tissues in which the receptor is expressed, such as disclosed herein.

[0114] The receptor polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the receptor protein.

[0115] Determining the ability of the test compound to interact with the polypeptide can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the polypeptide.

[0116] The polypeptides can be used to identify compounds that modulate receptor activity. Such compounds, for example, can increase or decrease affinity or rate of binding to a known ligand, compete with ligand for binding to the receptor, or displace ligand bound to the receptor. Both 14275 protein and appropriate variants and fragments can be used in high

throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

[0117] The receptor polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a target molecule that normally interacts with the receptor protein. The target can be ligand or a component of the signal pathway with which the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

[0118] Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

[0119] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[0120] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 97:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra*).

[0121] Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

[0122] Candidate compounds further include lysophospholipids, phospholipids, glycerophospholipids, sphingolipids, and lysosphingolipids. They can be related to natural ligands such as ceramide, sphingosine, S1P, LPA, cyclic LPA, cycosine, dihydrosphingosine, lysophosphatidyl-choline, lysophosphatidyl-ethanolamine, lysophosphatidyl serine, and lysosphingomyelin (sphingosyl-phosphorylcholine).

[0123] One candidate compound is a soluble full-length receptor or fragment that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

[0124] The invention provides other end points to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the receptor protein, or a receptor protein target, could also be measured.

[0125] Targets in signaling include any of the intermediates in lipid-mediated GPCR transduction including adenylyl cyclase, cAMP, receptor-G protein complex, G protein subunit disassociation, MAPK activation, activated Ras, P13K γ , activated tyrosine kinases, Rho-activated Ser/Thr kinases, and phosphorylated MLC.

[0126] Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemicals or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these end point assay targets, and other functions known to those of ordinary skill in the art.

[0127] Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain or part thereof, the region spanning the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops, and the carboxy terminal intracellular domain or part can be replaced by heterologous domains or parts thereof. For example, a G-protein-binding region can be used that interacts with a different G-protein than that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation.

Alternatively, one or more of the transmembrane segments or loops can be replaced with one or more of the transmembrane segments or loops specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain or a part thereof and/or other ligand-binding regions could be replaced by a domain or part thereof and/or other ligand-binding regions binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by

a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

[0128] The receptor polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

[0129] To perform cell free drug screening assays, it is desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

[0130] Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/14275 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount

of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0131] Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells that express the 14275 receptor protein, such as in mobilized peripheral blood CD34⁺ cells, promyelocytic leukemia cells, CD34⁺ mobilized bone marrow cells, CD8⁺T lymphocytes, CD34⁺ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, CD34⁺ mobilized peripheral blood cells, and CD4⁺T lymphocytes. These methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

[0132] The polypeptides are thus useful for treating a receptor -associated disorder characterized by aberrant expression or activity of a receptor protein. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves administering a protein as therapy to compensate for reduced or aberrant expression or activity of the protein.

[0133] Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example of such a situation, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example of such a situation, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

[0134] In yet another aspect of the invention, the proteins of the invention can be used as “bait proteins” in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

[0135] The receptor polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the receptor protein, especially in mobilized peripheral blood CD34⁺ cells, in promyelocytic leukemia cells, CD34⁺ mobilized bone marrow cells, CD8⁺T lymphocytes, CD34⁺ adult bone marrow cells, lymph node, leukocytes from G-CSF treated donors, CD34⁺ mobilized peripheral blood cells, and CD4⁺T lymphocytes. Accordingly, methods are provided for detecting the presence, or levels of, the receptor protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected.

[0136] One agent for detecting receptor protein is an antibody capable of selectively binding to receptor protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

[0137] The receptor protein also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant receptor protein. Thus, receptor protein can be isolated from a biological sample, assayed for the presence of a genetic mutation that results in aberrant receptor protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

[0138] *In vitro* techniques for detection of receptor protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-receptor antibody. For example, the antibody can

be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of a receptor protein expressed in a subject and methods which detect fragments of a receptor protein in a sample.

[0139] The receptor polypeptides are also useful in pharmacogenomic analysis. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

[0140] The receptor polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or receptor activity can be monitored over the course of treatment using the receptor polypeptides as an end-point target.

[0141] The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a specified protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

[0142] The receptor polypeptides are also useful for treating a receptor-associated disorder. Accordingly, methods for treatment include the use of soluble receptor or fragments of the receptor protein that compete for ligand binding. These receptors or fragments can have a higher affinity for the ligand so as to provide effective competition.

Antibodies

[0143] The invention also provides antibodies that selectively bind to the 14275 receptor protein and its variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially homologous with the receptor protein. These other proteins share homology with a fragment or domain of the receptor protein. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the receptor protein is still selective.

[0144] Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or F(ab')₂) can be used.

[0145] Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ¹²⁵I, ¹³¹I, ³⁵S or ³H.

[0146] To generate antibodies, an isolated receptor polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index can be found. Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire receptor or portions of the receptor, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against

specific functional sites, such as the site of ligand-binding, the site of G protein coupling, or sites that are phosphorylated, glycosylated, or myristoylated.

[0147] An antigenic fragment will typically comprise at least 6, 9, or 12 contiguous amino acid residues. The antigenic peptide can comprise a contiguous sequence of at least 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments which may be disclosed prior to the invention.

[0148] An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

Antibody Uses

[0149] The antibodies can be used to isolate a receptor protein by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural receptor protein from cells and recombinantly produced receptor protein expressed in host cells.

[0150] The antibodies are useful to detect the presence of receptor protein in cells or tissues to determine the pattern of expression of the receptor among various tissues in an organism and over the course of normal development.

[0151] The antibodies can be used to detect receptor protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

[0152] The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

[0153] Antibody detection of circulating fragments of the full length receptor protein can be used to identify receptor turnover.

[0154] Further, the antibodies can be used to assess receptor expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to receptor function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the receptor protein, the antibody can be prepared against the normal receptor protein. If a disorder is characterized by a specific mutation in the receptor protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant receptor protein.

[0155] The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole receptor or portions of the receptor, for example, portions of the amino terminal extracellular domain or extracellular loops.

[0156] The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting receptor expression level or the presence of aberrant receptors and aberrant tissue distribution or developmental expression, antibodies directed against the receptor or relevant fragments can be used to monitor therapeutic efficacy.

[0157] Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

[0158] Additionally, antibodies are useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23 (10-11):983-985 (1996)); and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes affects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Thus, antibodies prepared against polymorphic receptor proteins can be used to identify individuals that require modified treatment modalities.

[0159] The antibodies are also useful as diagnostic tools as an immunological marker for aberrant receptor protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

[0160] The antibodies are also useful for tissue typing. Thus, where a specific receptor protein has been correlated with expression in a specific tissue, antibodies that are specific for this receptor protein can be used to identify a tissue type.

[0161] The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

[0162] The antibodies are also useful for inhibiting receptor function, for example, blocking ligand binding.

[0163] These uses can also be applied in a therapeutic context in which treatment involves inhibiting receptor function. An antibody can be used, for example, to block ligand binding. Antibodies can be prepared against specific fragments containing sites required for function or against intact receptor associated with a cell.

[0164] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806.

[0165] The invention also encompasses kits for using antibodies to detect the presence of a receptor protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting receptor protein in a biological sample; means for determining the amount of receptor protein in the sample; and means for comparing the amount of receptor protein in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect receptor protein.

Polynucleotides

[0166] The nucleotide sequence in SEQ ID NO:1 was obtained by sequencing the human full length cDNA. The specifically disclosed cDNA comprises the coding region, 5' and 3' untranslated sequences (SEQ ID NO:1). In one embodiment, the receptor nucleic acid comprises only the coding region.

[0167] The human 14275 receptor cDNA is approximately 1877 nucleotides in length and encodes a full length protein that is approximately 384 amino acid residues in length. The nucleic acid is expressed in: thymus, colon, spleen, and peripheral blood cells with lower expression in the lung, heart, small intestine, uterus, prostate, and placenta. Structural analysis of the amino acid sequence of SEQ ID NO:2 can be performed using a hydropathy plot. The plot would show the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain. As used herein, the term "transmembrane segment" refers to a structural amino acid motif which includes a hydrophobic helix that spans the plasma membrane. The entire transmembrane domain spans amino acids from about 51 to about 331. Seven segments span the membrane and there are three intracellular and three extracellular loops in the domain.

[0168] The invention provides isolated polynucleotides encoding a 14275 receptor protein. The term "14275 polynucleotide" or "14275 nucleic acid" refers to the sequence shown in SEQ ID NO:1. The term "receptor polynucleotide" or "receptor nucleic acid" further includes variants and fragments of the 14275 polynucleotide.

[0169] An "isolated" receptor nucleic acid is one that is separated from other nucleic acid present in the natural source of the receptor nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the receptor nucleic acid sequences.

[0170] Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

[0171] For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially)

DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0172] In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.

[0173] The receptor polynucleotides can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

[0174] The receptor polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

[0175] Receptor polynucleotides can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

[0176] One receptor nucleic acid comprises the nucleotide sequence shown in SEQ ID NO:1, corresponding to human prostate cDNA.

[0177] The invention further provides variant receptor polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:1 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1.

[0178] The invention also provides receptor nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

[0179] Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

[0180] Typically, variants have a substantial identity with a nucleic acid molecule of SEQ ID NO:1 and the complements thereof.

[0181] Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a receptor that is at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:1 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:1 or a fragment of the sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all GPCRs, all family I GPCRs, or all EDG receptors. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

[0182] As used herein, the term “hybridizes under stringent conditions” is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a receptor polypeptide at least 50-55%, 55% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 90%, at least about 95% or more identical to each other remain hybridized to one another. Such stringent conditions are

known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated by reference. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. In another non-limiting example, nucleic acid molecules are allowed to hybridize in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more low stringency washes in 0.2X SSC/0.1% SDS at room temperature, or by one or more moderate stringency washes in 0.2X SSC/0.1% SDS at 42°C, or washed in 0.2X SSC/0.1% SDS at 65°C for high stringency. In one embodiment, an isolated receptor nucleic acid molecule that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[0183] As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

[0184] The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1 and the complements. In one embodiment, the nucleic acid consists of a portion of the nucleotide sequence of SEQ ID NO:1 and the complements. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

[0185] Furthermore, the invention provides polynucleotides that comprise a fragment of the full length receptor polynucleotides. The fragment can be single or double

stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

[0186] In one embodiment, an isolated receptor nucleic acid is at least 52 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1.

[0187] In another embodiment, an isolated receptor nucleic acid encodes the entire coding region from amino acid 1 to amino acid 384. In another embodiment the isolated receptor nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to amino acid 384. Fragments further include nucleic acid sequences encoding a portion of the amino acid sequence described herein and further including flanking nucleotide sequences at the 3' region. Other fragments include nucleotide sequences encoding the amino acid fragments described herein. Further fragments can include subfragments of the specific domains or sites described herein. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof.

[0188] Receptor nucleic acid fragments also include a fragment from around nucleotide 1 to around 483 and subfragments thereof greater than 7 nucleotides. Receptor nucleic acid fragments further include a nucleotide sequence from around 477 to around 1143 and subfragments thereof greater than 18 nucleotides. A further receptor nucleic acid fragment includes nucleic acid from around 1121 to around 1369 and subfragments thereof greater than 33 nucleotides. A further fragment is from about 1387-1425 and subfragments thereof greater than 11 nucleotides. A further fragment is from about 1425 to the end of the sequence and subfragments thereof greater than 7 nucleotides. In these embodiments, the nucleic acid can be at least 17, 20, 30, 40, 50, 100, 250, or 500 nucleotides in length or greater. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[0189] Receptor nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Receptor nucleic acid fragments include nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain including amino acid residues from 1 to about 50, a polypeptide comprising the region spanning the transmembrane domain (amino acid residues from about 51 to about 331), a polypeptide comprising the carboxy terminal intracellular domain (amino acid residues from about 332 to about 384), and a polypeptide

encoding the G--protein receptor signature (142-144 or surrounding amino acid residues from about 135 to about 150), nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites, protein kinase C phosphorylation sites, and casein kinase II phosphorylation sites and myristoylation sites.

[0190] Receptor nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a receptor nucleic acid could include sequences corresponding to the amino terminal extracellular domain and one transmembrane fragment. A person of ordinary skill in the art would be aware of the many permutations that are possible. Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

[0191] However, it is understood that a receptor fragment includes any nucleic acid sequence that does not include the entire gene.

[0192] The invention also provides receptor nucleic acid fragments that encode epitope bearing regions of the receptor proteins described herein.

[0193] The isolated receptor polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

[0194] For example, the coding region of a receptor gene can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of receptor genes.

[0195] A probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:1 sense or anti-sense strand or other receptor polynucleotides. A probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

Polynucleotide Uses

[0196] The nucleic acid sequences of the present invention can be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0197] The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. “Probes” are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science* 254:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of the nucleic acid sequence of SEQ ID NO:1 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

[0198] As used herein, the term “primer” refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term “primer site” refers to the area of the target DNA to which a primer hybridizes. The term “primer pair” refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

[0199] The receptor polynucleotides are useful for probes, primers, and in biological assays.

[0200] Where the polynucleotides are used to assess GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful.

In this case, even fragments that may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed to GPCR functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing receptor function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of receptor dysfunction, all fragments are encompassed including those which may have been known in the art.

[0201] The receptor polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO:2 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO:2 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO:2 was isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally controlled and therefore may be expressed in the same tissue at different points in the development of an organism.

[0202] The probe can correspond to any sequence along the entire length of the gene encoding the receptor. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

[0203] The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:2, or a fragment thereof, such as an oligonucleotide of at least 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

[0204] Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

[0205] The fragments are also useful to synthesize antisense molecules of desired length and sequence.

[0206] Antisense nucleic acids of the invention can be designed using the nucleotide sequence of SEQ ID NO:1, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring

nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

[0207] Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670. PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug

delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63, Mag *et al.* (1989) *Nucleic Acids Res.* 17:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119.

[0208] The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res.* 5:539-549).

[0209] The receptor polynucleotides are also useful as primers for PCR to amplify any given region of a receptor polynucleotide.

[0210] The receptor polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the receptor polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of receptor genes and gene products. For example, an endogenous receptor coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

[0211] The receptor polynucleotides are also useful as probes for determining the chromosomal positions of the receptor polynucleotides by means of *in situ* hybridization methods, such as FISH (for a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York), and PCR mapping of somatic cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

[0212] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[0213] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature* 325:783-787.

[0214] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

[0215] The receptor polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the receptors and their variants with respect to tissue distribution, for example whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously. The receptor polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

[0216] The receptor polynucleotides are also useful for constructing host cells expressing a part, or all, of the receptor polynucleotides and polypeptides.

[0217] The receptor polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the receptor polynucleotides and polypeptides.

[0218] The receptor polynucleotides are also useful for making vectors that express part, or all, of the receptor polypeptides.

[0219] The receptor polynucleotides are also useful as hybridization probes for determining the level of receptor nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, receptor nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA.

Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the receptor genes.

[0220] Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the receptor genes, as on extrachromosomal elements or as integrated into chromosomes in which the receptor gene is not normally found, for example as a homogeneously staining region.

[0221] These uses are relevant for diagnosis of disorders involving an increase or decrease in receptor expression relative to normal results, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder, such as, for example, the disorders disclosed in the Example herein.

[0222] Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of receptor nucleic acid, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

[0223] One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

[0224] *In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

[0225] Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a receptor protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated.

[0226] Nucleic acid expression assays are useful for drug screening to identify compounds that modulate receptor nucleic acid expression (e.g., antisense, polypeptides,

peptidomimetics, small molecules or other drugs). A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of receptor mRNA in the presence of the candidate compound is compared to the level of expression of receptor mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

[0227] Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals.

[0228] The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the receptor gene. The method typically includes assaying the ability of the compound to modulate the expression of the receptor nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired receptor nucleic acid expression.

[0229] The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the receptor nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

[0230] Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

[0231] The assay for receptor nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the receptor protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase. Thus, modulators of receptor gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of receptor mRNA in the presence of the candidate compound is compared to the level of expression of receptor mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When

expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

[0232] Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate receptor nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

[0233] Alternatively, a modulator for receptor nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the receptor nucleic acid expression.

[0234] The receptor polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the receptor gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

[0235] Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

[0236] The receptor polynucleotides are also useful in diagnostic assays for qualitative changes in receptor nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in receptor genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally occurring genetic mutations in the receptor gene and thereby determining whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement such as inversion or transposition, modification of genomic DNA such as aberrant methylation patterns or changes in gene copy number such as amplification. Detection of a mutated form of the receptor gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a receptor protein.

[0237] Mutations in the receptor gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

[0238] In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

[0239] It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

[0240] Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

[0241] Alternatively, mutations in a receptor gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

[0242] Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[0243] Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

[0244] Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

[0245] Furthermore, sequence differences between a mutant receptor gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

[0246] Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985); Cotton *et al.*, *Proc. Nat'l. Acad. Sci. USA* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *Proc. Nat'l. Acad. Sci. USA* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-

type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.* 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

[0247] In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[0248] The receptor polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the receptor gene that results in altered affinity for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the receptor. Accordingly, the receptor polynucleotides described herein can be used to assess the mutation content of the receptor gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

[0249] Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual.

Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

[0250] The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

[0251] The receptor polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[0252] The receptor polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

[0253] Furthermore, the receptor sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the receptor sequences described herein can be used to prepare two

PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

[0254] Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The receptor sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

[0255] If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

[0256] The receptor polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

[0257] The receptor polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique. Fragments are at least 12 bases.

[0258] The receptor polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of receptor probes

can be used to identify tissue by species and/or by organ type. In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

[0259] Alternatively, the receptor polynucleotides can be used directly to block transcription or translation of receptor gene expression by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable receptor gene expression, nucleic acids can be directly used for treatment.

[0260] The receptor polynucleotides are thus useful as antisense constructs to control receptor gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of receptor protein. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into receptor protein.

[0261] Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NO:1 which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO:1.

[0262] Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of receptor nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired receptor nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the receptor protein.

[0263] The receptor polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in receptor gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired receptor protein to treat the individual.

[0264] The invention also encompasses kits for detecting the presence of a receptor nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting receptor nucleic acid in a biological sample; means for determining the amount of receptor nucleic acid in the sample;

and means for comparing the amount of receptor nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect receptor mRNA or DNA.

Computer Readable Means

[0265] The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, “provided” refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

[0266] In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, “computer readable media” refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

[0267] As used herein, “recorded” refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

[0268] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of

the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[0269] By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[0270] As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[0271] As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

[0272] Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used

in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

[0273] For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem.* 17:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/host cells

[0274] The invention also provides vectors containing the receptor polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the receptor polynucleotides. When the vector is a nucleic acid molecule, the receptor polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

[0275] A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the receptor polynucleotides. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the receptor polynucleotides when the host cell replicates.

[0276] The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the receptor polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

[0277] Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the receptor polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the receptor polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

[0278] It is understood, however, that in some embodiments, transcription and/or translation of the receptor polynucleotides can occur in a cell-free system.

[0279] The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

[0280] In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

[0281] In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

[0282] A variety of expression vectors can be used to express a receptor polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

[0283] The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or

other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

[0284] The receptor polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

[0285] The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

[0286] As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the receptor polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.* (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

[0287] Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185:119-128, Academic Press, San Diego, California (1990)). Alternatively, the sequence of the polynucleotide of interest can be altered to

provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

[0288] The receptor polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

[0289] The receptor polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

[0290] In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B., *Nature* 329:840 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

[0291] The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the receptor polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook, J.; Fritsh, E. F.; and Maniatis, T., *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[0292] The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[0293] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such

as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

[0294] The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989))*.

[0295] Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the receptor polynucleotides can be introduced either alone or with other polynucleotides that are not related to the receptor polynucleotides such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the receptor polynucleotide vector.

[0296] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

[0297] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

[0298] While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

[0299] Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the receptor polypeptides or heterologous to these polypeptides.

[0300] Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

[0301] It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

[0302] It is understood that “host cells” and “recombinant host cells” refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0303] The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing receptor proteins or polypeptides that can be further purified to produce desired amounts of receptor protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

[0304] Host cells are also useful for conducting cell-based assays involving the receptor or receptor fragments. Thus, a recombinant host cell expressing a native receptor is useful to assay for compounds that stimulate or inhibit receptor function. This includes ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway.

[0305] Cell-based assays include NE-115 (Postma, cited above); *Xenopus* oocytes, especially for calcium efflux (An, *FEBS Lett.*, cited above) and Cl currents (Guo, cited above); Jurkat cells, especially for reporter assays using SRE-driven transcription (An, *FEBS Lett.*, cited above); HEK 293 and CHO cells, especially for reporter assays using SRE-driven transcription (An, *Biochem. Biophys. Res. Comm.*, cited above).

[0306] Host cells are also useful for identifying receptor mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant receptor (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native receptor.

[0307] Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a heterologous region spanning the entire transmembrane domain (or parts thereof) can be used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane domain (or parts thereof) compatible with the specific host cell is used to make the chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

[0308] Further, mutant receptors can be designed in which one or more of the various functions is engineered to be increased or decreased (i.e., ligand binding or G-protein binding) and used to augment or replace receptor proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant receptor or providing an aberrant receptor that provides a therapeutic result. In one embodiment, the cells provide receptors that are abnormally active.

[0309] In another embodiment, the cells provide receptors that are abnormally inactive. These receptors can compete with endogenous receptors in the individual.

[0310] In another embodiment, cells expressing receptors that cannot be activated, are introduced into an individual in order to compete with endogenous receptors for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by receptor activation would be beneficial.

[0311] Homologously recombinant host cells can also be produced that allow the *in situ* alteration of endogenous receptor polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the receptor polynucleotides or sequences proximal or distal to a receptor gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a receptor protein can be produced in a cell not normally producing it. Alternatively, increased expression of receptor protein can be effected in a cell normally producing the protein at a specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the receptor protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant receptor proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site.

[0312] In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered receptor gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell* 51:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous receptor gene is selected (see e.g., Li, E. *et al.*, *Cell* 69:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing

homologous recombination vectors and homologous recombinant animals are described further in Bradley, A., *Current Opinions in Biotechnology* 2:823-829 (1991); and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

[0313] The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a receptor protein and identifying and evaluating modulators of receptor protein activity.

[0314] Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

[0315] In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which receptor polynucleotide sequences have been introduced.

[0316] A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the receptor nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

[0317] Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the receptor protein to particular cells.

[0318] Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986)). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals

carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

[0319] In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (*Proc. Nat'l. Acad. Sci. USA* 89:6232-6236 (1992)). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.*, *Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

[0320] Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (*Nature* 385:810-813 (1997)); and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

[0321] Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, receptor activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* receptor function, including ligand interaction, the effect of specific mutant receptors on receptor function and ligand interaction, and the effect of

chimeric receptors. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more receptor functions.

[0322] In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the receptor protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the receptor protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

[0323] The receptor nucleic acid molecules, protein (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

[0324] As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following

components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

[0325] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0326] Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a receptor protein or anti-receptor antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for

the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0327] Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0328] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[0329] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[0330] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0331] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release

formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0332] It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

[0333] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al.*, *PNAS* 91:3054-3057 (1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[0334] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

[0335] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7

mg/kg, or 5 to 6 mg/kg body weight.

[0336] The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

[0337] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[0338] It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5

milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0339] This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

EXAMPLE

[0340] The expression of the 14275 receptor was studied using a Taqman analysis. This procedure involves RT-PCR, according to routine procedures. Extremely high expression was observed in mobilized peripheral blood CD34⁺ cells. High expression was observed in HL60 (promyelocytic leukemia) cell line, CD34⁻ mobilized bone marrow cells, CD8⁺T lymphocytes, CD34⁺ adult bone marrow cells, lymph node, and leukocytes. Significant expression was also observed in spleen, CD34⁻ mobilized peripheral blood cells, CD4⁺T lymphocytes, resting B lymphocytes, and resting peripheral blood mononuclear cells. There was higher expression in resting peripheral blood mononuclear cells than in PHA-activated peripheral blood mononuclear cells. Some expression was also observed in

granulocytes. In T lymphocytes (Th1 and Th2) stimulated with anti-CD3, expression decreased over time (6-48 hours).

[0341] In view of these expression data, expression of the 14275 receptor is relevant to immunological disorders and disorders involving inflammation. Such immune disorders include, but are not limited to, chronic inflammatory diseases and disorders, such as Crohn's disease, reactive arthritis, including Lyme disease, insulin-dependent diabetes, organ-specific autoimmunity, including multiple sclerosis, Hashimoto's thyroiditis and Grave's disease, contact dermatitis, psoriasis, graft rejection, graft versus host disease, sarcoidosis, atopic conditions, such as asthma and allergy, including allergic rhinitis, gastrointestinal allergies, including food allergies, eosinophilia, conjunctivitis, glomerular nephritis, certain pathogen susceptibilities such as helminthic (e.g., leishmaniasis), certain viral infections, including HIV, and bacterial infections, including tuberculosis and lepromatous leprosy.

[0342] Respiratory disorders include, but are not limited to, apnea, asthma, particularly bronchial asthma, berillium disease, bronchiectasis, bronchitis, bronchopneumonia, cystic fibrosis, diphtheria, dyspnea, emphysema, chronic obstructive pulmonary disease, allergic bronchopulmonary aspergillosis, pneumonia, acute pulmonary edema, pertussis, pharyngitis, atelectasis, Wegener's granulomatosis, Legionnaires disease, pleurisy, rheumatic fever, and sinusitis.

[0343] Further, in view of expression of the receptor in blood progenitor cells, expression of the receptor is relevant to blood cell formation and thus useful for treating and diagnosing anemia, neutropenia, and thrombocytopenia.

[0344] Analysis of expression was also done using standard Northern blotting procedures. Expression was observed in peripheral blood cells, such as T and B cells, spleen, lung, thymus, uterus, small intestine, colon, heart, prostate and placenta.

[0345] Accordingly, the methods disclosed herein, including but not limited to, methods for identifying agents that modulate the level or activity of the receptor nucleic acid or polypeptide in a cell, methods of screening a cell to identify an agent that modulates the level or activity of the polypeptide or nucleic acid in a cell, methods for identifying agents that interact with the polypeptide or nucleic acid in a cell, methods of screening a cell to identify an agent that interacts with the polypeptide or nucleic acid in a cell, methods for detecting the presence of the polypeptide or nucleic acid in a cell, methods for modulating the level or activity of the polypeptide or nucleic acid in a cell, and any method of diagnosis and treatment based on these generic methods, are particularly applicable for these disorders

and for and in the cell types in which the receptor is expressed or in which expression is abnormally low or absent.

[0346] In normal bone marrow, the myelocytic series (polymorphoneuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-30%. Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes), polymorphoneuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure 2-8) of *Immunology, Immunopathology and Immunity*, Fifth Edition, Sell *et al.* Simon and Schuster (1996), incorporated by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following: diseases involving hematopoietic stem cells; committed lymphoid progenitor cells; lymphoid cells including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia, monocytic; [leukemias are encompassed with and without differentiation]; chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute myelocytic leukemia, hematologic malignancies of monocyte-macrophage lineage, such as juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell

carcinoma; sarcoma, including kaposi's sarcoma; fibroadenoma and phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium; demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

[0347] Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

[0348] Disorders involving red cells include, but are not limited to, anemias, such as hemolytic anemias, including hereditary spherocytosis, hemolytic disease due to erythrocyte enzyme defects: glucose-6-phosphate dehydrogenase deficiency, sickle cell disease, thalassemia syndromes, paroxysmal nocturnal hemoglobinuria, immunohemolytic anemia, and hemolytic anemia resulting from trauma to red cells; and anemias of diminished erythropoiesis, including megaloblastic anemias, such as anemias of vitamin B12 deficiency: pernicious anemia, and anemia of folate deficiency, iron deficiency anemia, anemia of chronic disease, aplastic anemia, pure red cell aplasia, and other forms of marrow failure.

[0349] Disorders involving B-cells include, but are not limited to precursor B-cell neoplasms, such as lymphoblastic leukemia/lymphoma. Peripheral B-cell neoplasms include, but are not limited to, chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, diffuse large B-cell lymphoma, Burkitt lymphoma, plasma cell neoplasms, multiple myeloma, and related entities, lymphoplasmacytic lymphoma (Waldenström

macroglobulinemia), mantle cell lymphoma, marginal zone lymphoma (MALToma), and hairy cell leukemia.

[0350] Disorders related to reduced platelet number (thrombocytopenia) include idiopathic thrombocytopenic purpura, including acute idiopathic thrombocytopenic purpura, drug-induced thrombocytopenia, HIV-associated thrombocytopenia, and thrombotic microangiopathies: thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome.

[0351] Disorders involving precursor T-cell neoplasms include precursor T lymphoblastic leukemia/lymphoma. Disorders involving peripheral T-cell and natural killer cell neoplasms include T-cell chronic lymphocytic leukemia, large granular lymphocytic leukemia, mycosis fungoides and Sézary syndrome, peripheral T-cell lymphoma, unspecified, angioimmunoblastic T-cell lymphoma, angiocentric lymphoma (NK/T-cell lymphoma^{4a}), intestinal T-cell lymphoma, adult T-cell leukemia/lymphoma, and anaplastic large cell lymphoma.

[0352] Disorders involving the spleen include, but are not limited to, splenomegaly, including nonspecific acute splenitis, congestive splenomegaly, and splenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly include infections, such as nonspecific splenitis, infectious mononucleosis, tuberculosis, typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-azar, trypanosomiasis, schistosomiasis, leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic anemias, and thrombocytopenic purpura; immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

[0353] Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and

bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

[0354] Disorders involving the colon include, but are not limited to, congenital anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery, infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory disorders, including parasites and protozoa, acquired immunodeficiency syndrome, transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

[0355] Disorders involving the uterus and endometrium include, but are not limited to, endometrial histology in the menstrual cycle; functional endometrial disorders, such as anovulatory cycle, inadequate luteal phase, oral contraceptives and induced endometrial changes, and menopausal and postmenopausal changes; inflammations, such as chronic endometritis; adenomyosis; endometriosis; endometrial polyps; endometrial hyperplasia; malignant tumors, such as carcinoma of the endometrium; mixed Müllerian and mesenchymal tumors, such as malignant mixed Müllerian tumors; tumors of the myometrium, including leiomyomas, leiomyosarcomas, and endometrial stromal tumors.

[0356] Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

[0357] Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lymphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

[0358] Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

II. 54420, A NOVEL HUMAN CALCIUM CHANNEL

Background of the Invention

[0359] Calcium signaling has been implicated in the regulation of a variety of cellular responses, such as growth and differentiation. There are two general methods by which intracellular concentrations of calcium ions may be increased: calcium ions may be brought into the cell from the extracellular milieu through the use of specific channels in the cellular membrane, or calcium ions may be freed from intracellular stores, again being transported by specific membrane channels in the storage organelle. In the situation in which the intracellular stores of calcium have been depleted, a specific type of calcium channel, termed a 'capacitative calcium channel' or a 'store-operated calcium channel' (SOC), is activated in the plasma membrane to import calcium ions from the extracellular environment to the cytosol (for review, see Putney and McKay (1999) *BioEssays* 21:38-46).

[0360] Members of the capacitative calcium channel family include the calcium release-activated calcium current (CRAC) (Hoth and Penner (1992) *Nature* 355: 353-355), calcium release-activated nonselective cation current (CRANC) (Krause *et al.* (1996) *J. Biol. Chem.* 271: 32523-32528), and the transient receptor potential (TRP) proteins. There is no single electrophysiological profile characteristic of the family; rather, a wide array of single channel conductances, cation selectivity, and current properties have been observed for different specific channels. Further, in several instances it has been demonstrated that homo- or heteropolymerization of the channel molecule may occur, further changing the channel properties from that of the single molecule. In general, though, these channels function similarly, in that they are calcium ion-permeable cation channels which become activated upon stimulation of phospholipase C β by a G protein-coupled receptor. Depletion of intracellular calcium stores activate these channels by a mechanism which is yet undefined, but which has been demonstrated to involve a diffusible factor using studies in which calcium stores were artificially depleted (*e.g.*, by the introduction of chelators into the cell, by activating phospholipase C γ , or by inhibiting those enzymes responsible for pumping

calcium ions into the stores or those enzymes responsible for maintaining resting intracellular calcium ion concentrations) (Putney, J.W., (1986) *Cell Calcium* 7: 1-12; Putney, J.W. (1990) *Cell Calcium* 11:611-624).

[0361] The TRP channel family is one of the best characterized members of the capacitative calcium channel group. These channels include transient receptor potential protein and homologues thereof (to date, seven homologs and splice variants have been identified in a variety of organisms), the vanilloid receptor subtype I (also known as the capsaicin receptor), stretch-inhibitable non-selective cation channel (SIC), olfactory, mechanosensitive channel, insulin-like growth factor I-regulated calcium channel, and vitamin D-responsive apical, epithelial calcium channel (ECaC), melastatin, and the polycystic kidney disease protein family (see, *e.g.*, Montell and Rubin (1989) *Neuron* 2:1313-1323; Caterina *et al.* (1997) *Nature* 389: 816-824; Suzuki *et al.* (1999) *J. Biol. Chem.* 274: 6330-6335; Kiselyov *et al.* (1998) *Nature* 396: 478-482; Hoenderop *et al.* (1999) *J. Biol. Chem.* 274: 8375-8378; and Chen *et al.* (1999) *Nature* 401(6751): 383-6). Each of these molecules is 700 or more amino acids in length (TRP and TRP homologs have 1300 or more amino acid residues), and shares certain conserved structural features. Predominant among these structural features are six transmembrane domains, with an additional hydrophobic loop present between the fifth and sixth transmembrane domains. It is believed that this loop is integral to the activity of the pore of the channel formed upon membrane insertion (Hardie and Minke (1993) *Trends Neurosci* 16: 371-376). TRP channel proteins also include one or more ankyrin domains and frequently display a proline-rich region at the N-terminus. Although found in disparate tissues and organisms, members of the TRP channel protein family all serve to transduce signals by means of calcium entry into cells, particularly pain signals (see, *e.g.*, McClesky and Gold (1999) *Annu. Rev. Physiol.* 61: 835-856), light signals (Hardie and Minke, *supra*), or olfactory signals (Colbert *et al.* (1997) *J. Neurosci* 17(21): 8259-8269). Thus, this family of molecules may play important roles in sensory signal transduction in general.

Summary of the Invention

[0362] The present invention is based, at least in part, on the discovery of novel transient receptor potential (TRP) family members, referred to herein as TRP-like calcium channel or TLCC-2 nucleic acid and protein molecules. The TLCC-2 molecules of the present invention are useful as targets for developing modulating agents to regulate a variety

of cellular processes, including membrane excitability, neurite outgrowth and synaptogenesis, signal transduction, cell proliferation, growth, differentiation, and migration, and nociception. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding TLCC-2 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of TLCC-2-encoding nucleic acids.

[0363] In one embodiment, a TLCC-2 nucleic acid molecule of the invention is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the nucleotide sequence (*e.g.*, to the entire length of the nucleotide sequence) shown in SEQ ID NO:4 or 6, or a complement thereof.

[0364] In a preferred embodiment, the isolated nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:4 or 6, or a complement thereof. In another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 1-140 of SEQ ID NO:4. In yet another embodiment, the nucleic acid molecule includes SEQ ID NO:6 and nucleotides 1884-2095 of SEQ ID NO:4. In another preferred embodiment, the nucleic acid molecule consists of the nucleotide sequence shown in SEQ ID NO:4 or 6. In another preferred embodiment, the nucleic acid molecule includes a fragment of at least 706 nucleotides (*e.g.*, 706 contiguous nucleotides) of the nucleotide sequence of SEQ ID NO: 4 or 6, or a complement thereof.

[0365] In another embodiment, a TLCC-2 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:5. In a preferred embodiment, a TLCC-2 nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the entire length of the amino acid sequence of SEQ ID NO:5.

[0366] In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of human TLCC-2. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO:5. In yet another preferred embodiment, the nucleic acid molecule is at least 706 nucleotides in length. In a further preferred embodiment, the nucleic acid molecule is at least 706 nucleotides in length and encodes a protein having a TLCC-2 activity (as described herein).

[0367] Another embodiment of the invention features nucleic acid molecules, preferably TLCC-2 nucleic acid molecules, which specifically detect TLCC-2 nucleic acid molecules relative to nucleic acid molecules encoding non-TLCC-2 proteins. For example, in one embodiment, such a nucleic acid molecule is at least 706, 706-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1070, 1070-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1750, 1750-1800, or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:4, or a complement thereof.

[0368] In preferred embodiments, the nucleic acid molecules are at least 15 (*e.g.*, 15 contiguous) nucleotides in length and hybridize under stringent conditions to nucleotides 1-27 of SEQ ID NO:4.

[0369] In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:5, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:4 or 6 under stringent conditions.

[0370] Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a TLCC-2 nucleic acid molecule, *e.g.*, the coding strand of a TLCC-2 nucleic acid molecule.

[0371] Another aspect of the invention provides a vector comprising a TLCC-2 nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. In yet another embodiment, the invention provides a host cell containing a nucleic acid molecule of the invention. The invention also provides a method for producing a protein, preferably a TLCC-2 protein, by culturing in a suitable medium, a host cell, *e.g.*, a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

[0372] Another aspect of this invention features isolated or recombinant TLCC-2 proteins and polypeptides. In one embodiment, an isolated TLCC-2 protein has one or more of the following domains: a transmembrane domain, a pore domain, and a proline rich domain. In a preferred embodiment, a TLCC-2 protein includes at least one or more of the following domains: a transmembrane domain, a pore domain, and a proline rich domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%,

90%, 95%, 96%, 97%, 98%, 99% or more identical to the amino acid sequence of SEQ ID NO:5. In another preferred embodiment, a TLCC-2 protein includes at least one transmembrane domain and has a TLCC-2 activity (as described herein).

[0373] In yet another preferred embodiment, a TLCC-2 protein includes one or more of the following domains: a transmembrane domain, a pore domain, and a proline rich domain, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4 or 6.

[0374] In another embodiment, the invention features fragments of the protein having the amino acid sequence of SEQ ID NO:5, wherein the fragment comprises at least 18 or more amino acids (*e.g.*, contiguous amino acids) of the amino acid sequence of SEQ ID NO:5. In another embodiment, a TLCC-2 protein has the amino acid sequence of SEQ ID NO:5.

[0375] In another embodiment, the invention features a TLCC-2 protein which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NO:4 or 6, or a complement thereof. This invention further features a TLCC-2 protein, which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4 or 6, or a complement thereof.

[0376] The proteins of the present invention or portions thereof, *e.g.*, biologically active portions thereof, can be operatively linked to a non-TLCC-2 polypeptide (*e.g.*, heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably TLCC-2 proteins. In addition, the TLCC-2 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

[0377] In another aspect, the present invention provides a method for detecting the presence of a TLCC-2 nucleic acid molecule, protein, or polypeptide in a biological sample by contacting the biological sample with an agent capable of detecting a TLCC-2 nucleic acid molecule, protein, or polypeptide such that the presence of a TLCC-2 nucleic acid molecule, protein or polypeptide is detected in the biological sample.

[0378] In another aspect, the present invention provides a method for detecting the presence of TLCC-2 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of TLCC-2 activity such that the presence of TLCC-2 activity is detected in the biological sample.

[0379] In another aspect, the invention provides a method for modulating TLCC-2 activity comprising contacting a cell capable of expressing TLCC-2 with an agent that modulates TLCC-2 activity such that TLCC-2 activity in the cell is modulated. In one embodiment, the agent inhibits TLCC-2 activity. In another embodiment, the agent stimulates TLCC-2 activity. In one embodiment, the agent is an antibody that specifically binds to a TLCC-2 protein. In another embodiment, the agent modulates expression of TLCC-2 by modulating transcription of a TLCC-2 gene or translation of a TLCC-2 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a TLCC-2 mRNA or a TLCC-2 gene.

[0380] In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant or unwanted TLCC-2 protein or nucleic acid expression or activity by administering an agent which is a TLCC-2 modulator to the subject. In one embodiment, the TLCC-2 modulator is a TLCC-2 protein. In another embodiment the TLCC-2 modulator is a TLCC-2 nucleic acid molecule. In yet another embodiment, the TLCC-2 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant or unwanted TLCC-2 protein or nucleic acid expression is a CNS disorder, such as a neurodegenerative disorder, *e.g.*, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfeldt disease, or AIDS related dementia, familial infantile convulsions, paroxysmal choreoathetosis; a disorder of the conveyance of sensory impulses from the periphery to the brain and/or conductance of motor impulses from the brain to the periphery; a psychiatric disorder (*e.g.*, depression, schizophrenic disorders, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; or a learning or memory disorder (*e.g.*, amnesia or age-related memory loss; or is a neurological disorder (*e.g.*, migraine).

[0381] In another embodiment, the disorder characterized by aberrant or unwanted TLCC-2 activity is a pain disorder, or a disorder characterized by misregulated pain signaling mechanisms.

[0382] In another embodiment, the disorder characterized by aberrant or unwanted TLCC-2 activity is a cell proliferation, growth, differentiation, or migration disorder.

[0383] The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a TLCC-2 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a TLCC-2 protein, wherein a wild-type form of the gene encodes a protein with a TLCC-2 activity.

[0384] In another aspect the invention provides methods for identifying a compound that binds to or modulates the activity of a TLCC-2 protein, by providing an indicator composition comprising a TLCC-2 protein having TLCC-2 activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on TLCC-2 activity in the indicator composition to identify a compound that modulates the activity of a TLCC-2 protein.

[0385] Other features and advantages of the invention will be apparent from the following detailed description and claims.

Detailed Description of the Invention

[0386] The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as "TRP-like calcium channel" or "TLCC-2" nucleic acid and protein molecules, which are novel members of the ion channel, *e.g.*, calcium channel, family. These novel molecules are capable of, for example, modulating an ion-channel mediated activity (*e.g.*, a calcium channel-mediated activity) in a cell, *e.g.*, a neuronal, muscle (*e.g.*, cardiac muscle), or liver cell.

[0387] As used herein, an "ion channel" includes a protein or polypeptide which is involved in receiving, conducting, and transmitting signals in an electrically excitable cell, *e.g.*, a neuronal or muscle cell. Ion channels include calcium channels, potassium channels, and sodium channels. As used herein, a "calcium channel" includes a protein or polypeptide which is involved in receiving, conducting, and transmitting calcium ion-based signals in an electrically excitable cell. Calcium channels are calcium ion selective, and can determine membrane excitability (the ability of, for example, a neuronal cell to respond to a stimulus and to convert it into a sensory impulse). Calcium channels can also influence the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of

excitation. Calcium channels are typically expressed in electrically excitable cells, *e.g.*, neuronal cells, and may form heteromultimeric structures (*e.g.*, composed of more than one type of subunit). Calcium channels may also be found in non-excitable cells (*e.g.*, adipose cells or liver cells), where they may play a role in, *e.g.*, signal transduction. Examples of calcium channels include the low-voltage-gated channels and the high-voltage-gated channels. Calcium channels are described in, for example, Davila *et al.* (1999) *Annals New York Academy of Sciences* 868:102-17 and McEnery, M.W. *et al.* (1998) *J. Bioenergetics and Biomembranes* 30(4): 409-418, the contents of which are incorporated herein by reference. As the TLCC-2 molecules of the present invention may modulate ion channel mediated activities (*e.g.*, calcium channel mediated activities), they may be useful for developing novel diagnostic and therapeutic agents for ion channel associated disorders (*e.g.*, calcium channel associated disorders).

[0388] As used herein, an “ion channel associated disorder” includes a disorder, disease or condition which is characterized by a misregulation of ion channel mediated activity. For example, a “calcium channel associated disorder” includes a disorder, disease or condition which is characterized by a misregulation of calcium channel mediated activity. Ion channel associated disorders, *e.g.*, calcium channel associated disorders, include CNS disorders, such as cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, senile dementia, Huntington's disease, Gilles de la Tourette's syndrome, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfeldt disease, or AIDS related dementia; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss, attention deficit disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, *e.g.*, migraine and obesity. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

[0389] Ion channel associated disorders, *e.g.*, calcium channel disorders, also include pain disorders. Pain disorders include those that affect pain signaling mechanisms. As used herein, the term "pain signaling mechanisms" includes the cellular mechanisms involved in the development and regulation of pain, *e.g.*, pain elicited by noxious chemical, mechanical, or thermal stimuli, in a subject, *e.g.*, a mammal such as a human. In mammals, the initial detection of noxious chemical, mechanical, or thermal stimuli, a process referred to as "nociception", occurs predominantly at the peripheral terminals of specialized, small diameter sensory neurons. These sensory neurons transmit the information to the central nervous system, evoking a perception of pain or discomfort and initiating appropriate protective reflexes. The TLCC-2 molecules of the present invention may be present on these sensory neurons and, thus, may be involved in detecting these noxious chemical, mechanical, or thermal stimuli and transducing this information into membrane depolarization events. Thus, the TLCC-2 molecules by participating in pain signaling mechanisms, may modulate pain elicitation and act as targets for developing novel diagnostic targets and therapeutic agents to control pain.

[0390] Ion channel associated disorders, *e.g.*, calcium channel disorders, also include cellular proliferation, growth, differentiation, or migration disorders. Cellular proliferation, growth, differentiation, or migration disorders include those disorders that affect cell proliferation, growth, differentiation, or migration processes. As used herein, a "cellular proliferation, growth, differentiation, or migration process" is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. The TLCC-2 molecules of the present invention are involved in signal transduction mechanisms, which are known to be involved in cellular growth, differentiation, and migration processes. Thus, the TLCC-2 molecules may modulate cellular growth, differentiation, or migration, and may play a role in disorders characterized by aberrantly regulated growth, differentiation, or migration. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; neuronal deficiencies resulting from impaired neural induction and patterning; hepatic disorders; cardiovascular disorders; and hematopoietic and/or myeloproliferative disorders.

[0391] As used herein, an "ion channel mediated activity" includes an activity which involves an ion channel, *e.g.*, an ion channel in a neuronal cell, a muscular cell, or a liver

cell, associated with receiving, conducting, and transmitting signals, in, for example, the nervous system. Ion channel mediated activities (*e.g.*, calcium channel mediated activities) include release of neurotransmitters or second messenger molecules (*e.g.*, dopamine or norepinephrine), from cells, *e.g.*, neuronal cells; modulation of resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation; participation in signal transduction pathways, and modulation of processes such as integration of sub-threshold synaptic responses and the conductance of back-propagating action potentials in, for example, neuronal cells (*e.g.*, changes in those action potentials resulting in a morphological or differentiative response in the cell).

[0392] The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, monkey proteins. Members of a family may also have common functional characteristics.

[0393] For example, the family of TLCC-2 proteins comprise at least one "transmembrane domain" and preferably six transmembrane domains. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 20-45 amino acid residues in length which spans the plasma membrane. More preferably, a transmembrane domain includes about at least 20, 25, 30, 35, 40, or 45 amino acid residues and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an alpha-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, Zagotta W.N. *et al*, (1996) *Annual Rev. Neurosci.* 19: 235-263, the contents of which are incorporated herein by reference. Amino acid residues 70-86, 299-317, 354-371, 385-416, 428-447, and 497-521 of the TLCC-2 protein comprise transmembrane domains. Accordingly, TLCC-2 proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a transmembrane domain of human TLCC-2 are within the scope of the invention.

[0394] In another embodiment, a TLCC-2 molecule of the present invention is identified based on the presence of at least one pore domain between the fifth and sixth transmembrane domains. As used herein, the term “pore domain” includes an overall hydrophobic amino acid sequence which is located between two transmembrane domains of a calcium channel protein, preferably transmembrane domains 5 and 6, and which is believed to be a major determinant of ion selectivity and channel activity in calcium channels. Pore domains are described, for example in Vannier *et al.* (1998) *J. Biol. Chem.* 273: 8675-8679 and Phillips, A. M. *et al.* (1992) *Neuron* 8, 631-642, the contents of which are incorporated herein by reference. TLCC-2 molecules having at least one pore domain are within the scope of the invention. A pore domain may be found in the human TLCC-2 sequence (SEQ ID NO:5) at about residues 459-470.

[0395] In another embodiment, the Peptide Structure program in the GCG software package can be used to find structural, hydrophobic, and antigenic regions in the human TLCC-2 sequence (SEQ ID NO:5).

[0396] In another embodiment, a TLCC-2 molecule of the present invention is identified based on the presence of at least one N-glycosylation site. As used herein, the term “N-glycosylation site” includes an amino acid sequence of about 4 amino acid residues in length which serves as a glycosylation site. More preferably, an N-glycosylation site has the consensus sequence Asn-Xaa-Ser/Thr (where Xaa may be any amino acid) (SEQ ID NO:7). N-glycosylation sites are described in, for example, Prosite PDOC00001, available online through the Prosite database, the contents of which are incorporated herein by reference. Amino acid residues 159-162, 179-182, 220-223, and 230-233 of the TLCC-2 protein comprise N-glycosylation sites. Accordingly, TLCC-2 proteins having at least one N-glycosylation site are within the scope of the invention.

[0397] In another embodiment, a TLCC-2 molecules of the present invention is identified based on the presence of a “proline rich domain” in the protein or corresponding nucleic acid molecule. As used herein, the term “proline rich domain” includes an amino acid sequence of about 4-6 amino acid residues in length having the general sequence Xaa-Pro-Xaa-Xaa-Pro-Xaa (where Xaa can be any amino acid; SEQ ID NO:62). Proline rich domains are usually located in a helical structure and bind through hydrophobic interactions to SH3 domains. SH3 domains recognize proline rich domains in both forward and reverse orientations. Proline rich domains are described in, for example, Sattler, M. *et al.* (1998) *Leukemia* 12: 637-644, the contents of which are incorporated herein by reference. Residues

1-37 of the amino acid sequence of human TLCC-2 (SEQ ID NO:5) contain proline-rich domains.

[0398] In a preferred embodiment, the TLCC-2 molecules of the invention include at least one transmembrane domain, at least one N-glycosylation site, at least one pore domain, and at least one proline rich domain.

[0399] Isolated proteins of the present invention, preferably TLCC-2 proteins, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:5 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:4 or 6. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently identical.

[0400] As used interchangeably herein, an "TLCC-2 activity", "biological activity of TLCC-2" or "functional activity of TLCC-2", refers to an activity exerted by a TLCC-2 protein, polypeptide or nucleic acid molecule on a TLCC-2 responsive cell or tissue, or on a TLCC-2 protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a TLCC-2 activity is a direct activity, such as an association with a TLCC-2-target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which a TLCC-2 protein binds or interacts in nature, such that TLCC-2-mediated function is achieved. A TLCC-2 target molecule can be a non-TLCC-2 molecule or a TLCC-2 protein or polypeptide of the present invention. In an exemplary embodiment, a TLCC-2 target molecule is a TLCC-2 ligand, *e.g.*, a calcium channel ligand. Alternatively, a TLCC-2 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the TLCC-2 protein with a TLCC-2 ligand. The biological activities of TLCC-2 are described herein. For example, the TLCC-2 proteins of the present

invention can have one or more of the following activities: (1) modulate membrane excitability, (2) influence the resting potential of membranes, (3) modulate wave forms and frequencies of action potentials, (4) modulate thresholds of excitation, (5) modulate neurite outgrowth and synaptogenesis, (6) modulate signal transduction, and (7) participate in nociception.

[0401] Accordingly, another embodiment of the invention features isolated TLCC-2 proteins and polypeptides having a TLCC-2 activity. Preferred proteins are TLCC-2 proteins having at least one or more of the following domains: a transmembrane domain, an N-glycosylation site, a pore domain, and a proline rich domain, and, preferably, a TLCC-2 activity.

[0402] Additional preferred proteins have one or more of the following domains: a transmembrane domain, an N-glycosylation site, a pore domain, and a proline rich domain, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4 or 6.

[0403] The nucleotide sequence of the isolated human TLCC-2 cDNA and the predicted amino acid sequence of the human TLCC-2 polypeptide are shown in SEQ ID NOs:4 and 5, respectively.

[0404] The human TLCC-2 gene, which is approximately 2095 nucleotides in length, encodes a protein having a molecular weight of approximately 65.7 kD and which is approximately 580 amino acid residues in length.

[0405] Various aspects of the invention are described in further detail in the following subsections:

Isolated Nucleic Acid Molecules

[0406] One aspect of the invention pertains to isolated nucleic acid molecules that encode TLCC-2 proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify TLCC-2-encoding nucleic acid molecules (*e.g.*, TLCC-2 mRNA) and fragments for use as PCR primers for the amplification or mutation of TLCC-2 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using

nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0407] The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated TLCC-2 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0408] A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:4 or 6, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:4 or 6 as a hybridization probe, TLCC-2 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

[0409] Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:4 or 6 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:4 or 6.

[0410] A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to TLCC-2 nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

[0411] In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:4. The sequence of SEQ ID NO:4 corresponds to the human TLCC-2 cDNA. This cDNA comprises sequences encoding the human TLCC-2 protein (*i.e.*, "the coding region", from nucleotides 141-1883), as well as 5' untranslated sequences (nucleotides 1-140) and 3' untranslated sequences (nucleotides 1884-2095). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:4 (*e.g.*, nucleotides 141-1883, corresponding to SEQ ID NO:6).

[0412] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:4 or 6, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:4 or 6 is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:4 or 6, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:4 or 6, thereby forming a stable duplex.

[0413] In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the entire length of the nucleotide sequence shown in SEQ ID NO:4 or 6, or a portion of any of these nucleotide sequences.

[0414] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:4 or 6, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a TLCC-2 protein, *e.g.*, a biologically active portion of a TLCC-2 protein. The nucleotide sequence determined from the cloning of the TLCC-2 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other TLCC-2 family members, as well as TLCC-2 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, or 100 or more consecutive nucleotides of a sense sequence of SEQ ID NO:4 or 6, of an anti-sense sequence of SEQ ID NO:4 or 6, or of a naturally occurring allelic variant or mutant of SEQ ID NO:4 or 6. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is greater than 706, 706-750, 750-800,

800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1070, 1070-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1750, 1750-1800 or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:4.

[0415] Probes based on the TLCC-2 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a TLCC-2 protein, such as by measuring a level of a TLCC-2-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting TLCC-2 mRNA levels or determining whether a genomic TLCC-2 gene has been mutated or deleted.

[0416] A nucleic acid fragment encoding a "biologically active portion of a TLCC-2 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:4 or 6, which encodes a polypeptide having a TLCC-2 biological activity (the biological activities of the TLCC-2 proteins are described herein), expressing the encoded portion of the TLCC-2 protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the TLCC-2 protein.

[0417] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:4 or 6, due to degeneracy of the genetic code and thus encode the same TLCC-2 proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:4 or 6. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:5.

[0418] In addition to the TLCC-2 nucleotide sequences shown in SEQ ID NO:4 or 6, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the TLCC-2 proteins may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the TLCC-2 genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a TLCC-2 protein, preferably a mammalian TLCC-2 protein, and can further include non-coding regulatory sequences, and introns.

[0419] Allelic variants of human TLCC-2 include both functional and non-functional TLCC-2 proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the human TLCC-2 protein that maintain the ability to bind a TLCC-2 ligand or substrate and/or modulate membrane excitability or signal transduction. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:5, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein.

[0420] Non-functional allelic variants are naturally occurring amino acid sequence variants of the human TLCC-2 protein that do not have the ability to form functional calcium channels or to modulate membrane excitability. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:5, or a substitution, insertion or deletion in critical residues or critical regions.

[0421] The present invention further provides non-human orthologues of the human TLCC-2 proteins. Orthologues of the human TLCC-2 protein are proteins that are isolated from non- non-human organisms and possess the same TLCC-2 ligand binding and/or modulation of membrane excitation mechanisms of the human TLCC-2 protein. Orthologues of the human TLCC-2 protein can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:5.

[0422] Moreover, nucleic acid molecules encoding other TLCC-2 family members and, thus, which have a nucleotide sequence which differs from the TLCC-2 sequences of SEQ ID NO:4 or 6 are intended to be within the scope of the invention. For example, another TLCC-2 cDNA can be identified based on the nucleotide sequence of human TLCC-2. Moreover, nucleic acid molecules encoding TLCC-2 proteins from different species, and which, thus, have a nucleotide sequence which differs from the TLCC-2 sequences of SEQ ID NO:4 or 6 are intended to be within the scope of the invention. For example, a mouse TLCC-2 cDNA can be identified based on the nucleotide sequence of a human TLCC-2.

[0423] Nucleic acid molecules corresponding to natural allelic variants and homologues of the TLCC-2 cDNAs of the invention can be isolated based on their homology to the TLCC-2 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural

allelic variants and homologues of the TLCC-2 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the TLCC-2 gene.

[0424] Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4 or 6. In other embodiment, the nucleic acid is at least 706, 706-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1070, 1070-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1750, 1750-1800 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% identical to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:4 or 6 and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

[0425] In addition to naturally-occurring allelic variants of the TLCC-2 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:4 or 6, thereby leading to changes in the amino acid sequence of the encoded TLCC-2 proteins, without altering the functional ability of the TLCC-2 proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:4 or 6. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of TLCC-2 (*e.g.*, the sequence of SEQ ID NO:5) without altering the biological activity, whereas an "essential" amino acid residue is required for

biological activity. For example, amino acid residues that are conserved among the TLCC-2 proteins of the present invention, *e.g.*, those present in a transmembrane domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the TLCC-2 proteins of the present invention and other members of the TLCC-2 family are not likely to be amenable to alteration.

[0426] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding TLCC-2 proteins that contain changes in amino acid residues that are not essential for activity. Such TLCC-2 proteins differ in amino acid sequence from SEQ ID NO:5, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:5.

[0427] An isolated nucleic acid molecule encoding a TLCC-2 protein identical to the protein of SEQ ID NO:5, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:4 or 6, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:4 or 6 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a TLCC-2 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a TLCC-2 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for TLCC-2 biological activity to identify mutants that retain activity.

Following mutagenesis of SEQ ID NO:4 or 6, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[0428] In a preferred embodiment, a mutant TLCC-2 protein can be assayed for the ability to (1) modulate membrane excitability, (2) influence the resting potential of membranes, (3) modulate wave forms and frequencies of action potentials, (4) modulate thresholds of excitation, (5) modulate neurite outgrowth and synaptogenesis, (6) modulate signal transduction, and (7) participate in nociception.

[0429] In addition to the nucleic acid molecules encoding TLCC-2 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire TLCC-2 coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding TLCC-2. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human TLCC-2 corresponds to SEQ ID NO:6). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding TLCC-2. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

[0430] Given the coding strand sequences encoding TLCC-2 disclosed herein (*e.g.*, SEQ ID NO:6), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of TLCC-2 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of TLCC-2 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of TLCC-2 mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using

naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0431] The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a TLCC-2 protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense

nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[0432] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-O-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

[0433] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave TLCC-2 mRNA transcripts to thereby inhibit translation of TLCC-2 mRNA. A ribozyme having specificity for a TLCC-2-encoding nucleic acid can be designed based upon the nucleotide sequence of a TLCC-2 cDNA disclosed herein (*i.e.*, SEQ ID NO:4 or 6). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a TLCC-2-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, TLCC-2 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[0434] Alternatively, TLCC-2 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the TLCC-2 (*e.g.*, the TLCC-2 promoter and/or enhancers; *e.g.*, nucleotides 1-137 of SEQ ID NO:4) to form triple helical structures that prevent transcription of the TLCC-2 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

[0435] In yet another embodiment, the TLCC-2 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone

to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al. Proc. Natl. Acad. Sci.* 93: 14670-675.

[0436] PNAs of TLCC-2 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of TLCC-2 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

[0437] In another embodiment, PNAs of TLCC-2 can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of TLCC-2 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17:

5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

[0438] In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0439] Alternatively, the expression characteristics of an endogenous TLCC-2 gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous TLCC-2 gene. For example, an endogenous TLCC-2 gene which is normally "transcriptionally silent", *i.e.*, a TLCC-2 gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous TLCC-2 gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

[0440] A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous TLCC-2 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

Isolated TLCC-2 Proteins and Anti-TLCC-2 Antibodies

[0441] One aspect of the invention pertains to isolated TLCC-2 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as

immunogens to raise anti-TLCC-2 antibodies. In one embodiment, native TLCC-2 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, TLCC-2 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a TLCC-2 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

[0442] An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the TLCC-2 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of TLCC-2 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of TLCC-2 protein having less than about 30% (by dry weight) of non-TLCC-2 protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-TLCC-2 protein, still more preferably less than about 10% of non-TLCC-2 protein, and most preferably less than about 5% non-TLCC-2 protein. When the TLCC-2 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

[0443] The language "substantially free of chemical precursors or other chemicals" includes preparations of TLCC-2 protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of TLCC-2 protein having less than about 30% (by dry weight) of chemical precursors or non-TLCC-2 chemicals, more preferably less than about 20% chemical precursors or non-TLCC-2 chemicals, still more preferably less than about 10% chemical precursors or non-TLCC-2 chemicals, and most preferably less than about 5% chemical precursors or non-TLCC-2 chemicals.

[0444] As used herein, a "biologically active portion" of a TLCC-2 protein includes a fragment of a TLCC-2 protein which participates in an interaction between a TLCC-2 molecule and a non-TLCC-2 molecule. Biologically active portions of a TLCC-2 protein

include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the TLCC-2 protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:5, which include less amino acids than the full length TLCC-2 proteins, and exhibit at least one activity of a TLCC-2 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the TLCC-2 protein, *e.g.*, modulating membrane excitation mechanisms. A biologically active portion of a TLCC-2 protein can be a polypeptide which is, for example, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 274, 500, 525, 550, 575, 600, 625, 650, 675, or 700 or more amino acids in length. Biologically active portions of a TLCC-2 protein can be used as targets for developing agents which modulate a TLCC-2 mediated activity, *e.g.*, a membrane excitation mechanism.

[0445] In one embodiment, a biologically active portion of a TLCC-2 protein comprises at least one transmembrane domain. It is to be understood that a preferred biologically active portion of a TLCC-2 protein of the present invention comprises at least one or more of the following domains: a transmembrane domain, an N-glycosylation site, a pore domain, and a proline rich domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native TLCC-2 protein.

[0446] In a preferred embodiment, the TLCC-2 protein has an amino acid sequence shown in SEQ ID NO:5. In other embodiments, the TLCC-2 protein is substantially identical to SEQ ID NO:5, and retains the functional activity of the protein of SEQ ID NO:5, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the TLCC-2 protein is a protein which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:5.

[0447] To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%,

or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the TLCC-2 amino acid sequence of SEQ ID NO:5 having 580 amino acid residues, at least 50, preferably at least 100, more preferably at least 200, even more preferably at least 300, and even more preferably at least 400 or 500 or more amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0448] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*Comput. Appl. Biosci.*, 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0449] The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to TLCC-2 nucleic acid molecules of the invention. BLAST protein searches can be performed with the

XBLAST program, score = 100, wordlength = 3, and a Blosum62 matrix to obtain amino acid sequences homologous to TLCC-2 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used.

[0450] The invention also provides TLCC-2 chimeric or fusion proteins. As used herein, a TLCC-2 "chimeric protein" or "fusion protein" comprises a TLCC-2 polypeptide operatively linked to a non-TLCC-2 polypeptide. A "TLCC-2 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to TLCC-2, whereas a "non-TLCC-2 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the TLCC-2 protein, *e.g.*, a protein which is different from the TLCC-2 protein and which is derived from the same or a different organism. Within a TLCC-2 fusion protein the TLCC-2 polypeptide can correspond to all or a portion of a TLCC-2 protein. In a preferred embodiment, a TLCC-2 fusion protein comprises at least one biologically active portion of a TLCC-2 protein. In another preferred embodiment, a TLCC-2 fusion protein comprises at least two biologically active portions of a TLCC-2 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the TLCC-2 polypeptide and the non-TLCC-2 polypeptide are fused in-frame to each other. The non-TLCC-2 polypeptide can be fused to the N-terminus or C-terminus of the TLCC-2 polypeptide.

[0451] For example, in one embodiment, the fusion protein is a GST-TLCC-2 fusion protein in which the TLCC-2 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant TLCC-2.

[0452] In another embodiment, the fusion protein is a TLCC-2 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of TLCC-2 can be increased through the use of a heterologous signal sequence.

[0453] The TLCC-2 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The TLCC-2 fusion proteins can be used to affect the bioavailability of a TLCC-2 substrate. Use of TLCC-2 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a TLCC-2 protein; (ii)

mis-regulation of the TLCC-2 gene; and (iii) aberrant post-translational modification of a TLCC-2 protein.

[0454] Moreover, the TLCC-2-fusion proteins of the invention can be used as immunogens to produce anti-TLCC-2 antibodies in a subject, to purify TLCC-2 ligands and in screening assays to identify molecules which inhibit the interaction of TLCC-2 with a TLCC-2 substrate.

[0455] Preferably, a TLCC-2 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A TLCC-2-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the TLCC-2 protein.

[0456] The present invention also pertains to variants of the TLCC-2 proteins which function as either TLCC-2 agonists (mimetics) or as TLCC-2 antagonists. Variants of the TLCC-2 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a TLCC-2 protein. An agonist of the TLCC-2 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a TLCC-2 protein. An antagonist of a TLCC-2 protein can inhibit one or more of the activities of the naturally occurring form of the TLCC-2 protein by, for example, competitively modulating a TLCC-2-mediated activity of a TLCC-2 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the TLCC-2 protein.

[0457] In one embodiment, variants of a TLCC-2 protein which function as either TLCC-2 agonists (mimetics) or as TLCC-2 antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a TLCC-2 protein for TLCC-2 protein agonist or antagonist activity. In one embodiment, a variegated library of TLCC-2 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of TLCC-2 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential TLCC-2 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of TLCC-2 sequences therein. There are a variety of methods which can be used to produce libraries of potential TLCC-2 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential TLCC-2 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477).

[0458] In addition, libraries of fragments of a TLCC-2 protein coding sequence can be used to generate a variegated population of TLCC-2 fragments for screening and subsequent selection of variants of a TLCC-2 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a TLCC-2 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the TLCC-2 protein.

[0459] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of TLCC-2

proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify TLCC-2 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

[0460] In one embodiment, cell based assays can be exploited to analyze a variegated TLCC-2 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, an endothelial cell line, which ordinarily responds to TLCC-2 in a particular TLCC-2 substrate-dependent manner. The transfected cells are then contacted with TLCC-2 and the effect of expression of the mutant on signaling by the TLCC-2 substrate can be detected, *e.g.*, by monitoring intracellular calcium, IP3, or diacylglycerol concentration, phosphorylation profile of intracellular proteins, or the activity of a TLCC-2-regulated transcription factor. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the TLCC-2 substrate, and the individual clones further characterized.

[0461] An isolated TLCC-2 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind TLCC-2 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length TLCC-2 protein can be used or, alternatively, the invention provides antigenic peptide fragments of TLCC-2 for use as immunogens. The antigenic peptide of TLCC-2 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:5 and encompasses an epitope of TLCC-2 such that an antibody raised against the peptide forms a specific immune complex with TLCC-2. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[0462] Preferred epitopes encompassed by the antigenic peptide are regions of TLCC-2 that are located on the surface of the protein, *e.g.*, hydrophilic regions, as well as regions with high antigenicity.

[0463] A TLCC-2 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed TLCC-2 protein or a chemically synthesized TLCC-2 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic TLCC-2 preparation induces a polyclonal anti-TLCC-2 antibody response.

[0464] Accordingly, another aspect of the invention pertains to anti-TLCC-2 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as TLCC-2. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind TLCC-2. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of TLCC-2. A monoclonal antibody composition thus typically displays a single binding affinity for a particular TLCC-2 protein with which it immunoreacts.

[0465] Polyclonal anti-TLCC-2 antibodies can be prepared as described above by immunizing a suitable subject with a TLCC-2 immunogen. The anti-TLCC-2 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized TLCC-2. If desired, the antibody molecules directed against TLCC-2 can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-TLCC-2 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole

et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a TLCC-2 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds TLCC-2.

[0466] Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-TLCC-2 monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind TLCC-2, *e.g.*, using a standard ELISA assay.

[0467] Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-TLCC-2 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with TLCC-

2 to thereby isolate immunoglobulin library members that bind TLCC-2. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* *Nature* (1990) 348:552-554.

[0468] Additionally, recombinant anti-TLCC-2 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-

525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

[0469] An anti-TLCC-2 antibody (*e.g.*, monoclonal antibody) can be used to isolate TLCC-2 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-TLCC-2 antibody can facilitate the purification of natural TLCC-2 from cells and of recombinantly produced TLCC-2 expressed in host cells. Moreover, an anti-TLCC-2 antibody can be used to detect TLCC-2 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the TLCC-2 protein. Anti-TLCC-2 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Recombinant Expression Vectors and Host Cells

[0470] Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a TLCC-2 protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the

host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0471] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, TLCC-2 proteins, mutant forms of TLCC-2 proteins, fusion proteins, and the like).

[0472] The recombinant expression vectors of the invention can be designed for expression of TLCC-2 proteins in prokaryotic or eukaryotic cells. For example, TLCC-2 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further

in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0473] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0474] Purified fusion proteins can be utilized in TLCC-2 activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for TLCC-2 proteins, for example. In a preferred embodiment, a TLCC-2 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

[0475] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid *trp-lac* fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 *gn10-lac* fusion promoter mediated by a coexpressed viral RNA polymerase (T7 *gn1*). This viral polymerase is supplied by host

strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 *gn1* gene under the transcriptional control of the lacUV 5 promoter.

[0476] One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0477] In another embodiment, the TLCC-2 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).

[0478] Alternatively, TLCC-2 proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

[0479] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[0480] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-

specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddie (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[0481] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to TLCC-2 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

[0482] Another aspect of the invention pertains to host cells into which a TLCC-2 nucleic acid molecule of the invention is introduced, *e.g.*, a TLCC-2 nucleic acid molecule within a recombinant expression vector or a TLCC-2 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny

or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0483] A host cell can be any prokaryotic or eukaryotic cell. For example, a TLCC-2 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0484] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

[0485] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a TLCC-2 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0486] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a TLCC-2 protein. Accordingly, the invention further provides methods for producing a TLCC-2 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a TLCC-2 protein has been

introduced) in a suitable medium such that a TLCC-2 protein is produced. In another embodiment, the method further comprises isolating a TLCC-2 protein from the medium or the host cell.

[0487] The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which TLCC-2-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous TLCC-2 sequences have been introduced into their genome or homologous recombinant animals in which endogenous TLCC-2 sequences have been altered. Such animals are useful for studying the function and/or activity of a TLCC-2 and for identifying and/or evaluating modulators of TLCC-2 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous TLCC-2 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

[0488] A transgenic animal of the invention can be created by introducing a TLCC-2-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The TLCC-2 cDNA sequence of SEQ ID NO:4 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human TLCC-2 gene, such as a mouse or rat TLCC-2 gene, can be used as a transgene. Alternatively, a TLCC-2 gene homologue, such as another TLCC-2 family member, can be isolated based on hybridization to the TLCC-2 cDNA sequences of SEQ ID NO:4 or 6 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a TLCC-2 transgene to direct expression of a TLCC-2 protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a TLCC-2 transgene in its genome and/or expression of TLCC-2 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a TLCC-2 protein can further be bred to other transgenic animals carrying other transgenes.

[0489] To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a TLCC-2 gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the TLCC-2 gene. The TLCC-2 gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:6), but more preferably, is a non-human homologue of a human TLCC-2 gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:4). For example, a mouse TLCC-2 gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous TLCC-2 gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous TLCC-2 gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous TLCC-2 gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous TLCC-2 protein). In the homologous recombination nucleic acid molecule, the altered portion of the TLCC-2 gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the TLCC-2 gene to allow for homologous recombination to occur between the exogenous TLCC-2 gene carried by the homologous recombination nucleic acid molecule and an endogenous TLCC-2 gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking

TLCC-2 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced TLCC-2 gene has homologously recombined with the endogenous TLCC-2 gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then be injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

[0490] In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

[0491] Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmot, I. *et al.* (1997) *Nature* 385:810-813

and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

Pharmaceutical Compositions

[0492] The TLCC-2 nucleic acid molecules, fragments of TLCC-2 proteins, and anti-TLCC-2 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

[0493] A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or

sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0494] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0495] Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of a TLCC-2 protein or an anti-TLCC-2 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0496] Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral

therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0497] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0498] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[0499] The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0500] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be

prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0501] It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

[0502] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0503] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0504] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight,

preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0505] In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

[0506] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

[0507] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to

about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0508] Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

[0509] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A,

pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[0510] Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in *Controlled Drug Delivery* (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in *Monoclonal Antibodies '84: Biological And Clinical Applications*, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.*, 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0511] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[0512] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Uses and Methods of the Invention

[0513] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a TLCC-2 protein of the invention has one or more of the following activities: (1) modulates membrane excitability, (2) influences the resting potential of membranes, (3) modulates wave forms and frequencies of action potentials, (4) modulates thresholds of excitation, (5) modulates neurite outgrowth and synaptogenesis, (6) modulates signal transduction, and (7) participates in nociception.

[0514] The isolated nucleic acid molecules of the invention can be used, for example, to express TLCC-2 protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect TLCC-2 mRNA (*e.g.*, in a biological sample) or a genetic alteration in a TLCC-2 gene, and to modulate TLCC-2 activity, as described further below. The TLCC-2 proteins can be used to treat disorders characterized by insufficient or excessive production of a TLCC-2 substrate or production of TLCC-2 inhibitors. In addition, the TLCC-2 proteins can be used to screen for naturally occurring TLCC-2 substrates, to screen for drugs or compounds which modulate TLCC-2 activity, as well as to treat disorders characterized by insufficient or excessive production of TLCC-2 protein or production of TLCC-2 protein forms which have decreased, aberrant or unwanted activity compared to TLCC-2 wild type protein (*e.g.*, CNS disorders (such as neurodegenerative disorders), pain disorders, or disorders of cellular growth, differentiation, or migration. Moreover, the anti-TLCC-2 antibodies of the invention can be used to detect and isolate TLCC-2 proteins, to regulate the bioavailability of TLCC-2 proteins, and modulate TLCC-2 activity.

Screening Assays:

[0515] The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to TLCC-2 proteins, have a stimulatory or inhibitory effect on, for example, TLCC-2 expression or TLCC-2 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of TLCC-2 substrate.

[0516] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a TLCC-2 protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a TLCC-2 protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[0517] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

[0518] Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

[0519] In one embodiment, an assay is a cell-based assay in which a cell which expresses a TLCC-2 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate TLCC-2 activity is determined. Determining the ability of the test compound to modulate TLCC-2 activity can be accomplished by monitoring, for example, intracellular calcium, IP3, or diacylglycerol concentration, phosphorylation profile of intracellular proteins, or the activity of a TLCC-2-regulated transcription factor. The cell, for example, can be of mammalian origin, *e.g.*, a neuronal cell, or a liver cell.

[0520] The ability of the test compound to modulate TLCC-2 binding to a substrate or to bind to TLCC-2 can also be determined. Determining the ability of the test compound to modulate TLCC-2 binding to a substrate can be accomplished, for example, by coupling the TLCC-2 substrate with a radioisotope or enzymatic label such that binding of the TLCC-2 substrate to TLCC-2 can be determined by detecting the labeled TLCC-2 substrate in a complex. Alternatively, TLCC-2 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate TLCC-2 binding to a TLCC-2 substrate in a complex. Determining the ability of the test compound to bind TLCC-2 can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to TLCC-2 can be determined by detecting the labeled TLCC-2 compound in a complex. For example, compounds (*e.g.*, TLCC-2 substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[0521] It is also within the scope of this invention to determine the ability of a compound (*e.g.*, a TLCC-2 substrate) to interact with TLCC-2 without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with TLCC-2 without the labeling of either the compound or the TLCC-2. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and TLCC-2.

[0522] In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a TLCC-2 target molecule (*e.g.*, a TLCC-2 substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the TLCC-2 target molecule. Determining the ability of the test compound to modulate the activity of a TLCC-2 target molecule can be accomplished, for example, by determining the ability of the TLCC-2 protein to bind to or interact with the TLCC-2 target molecule.

[0523] Determining the ability of the TLCC-2 protein, or a biologically active fragment thereof, to bind to or interact with a TLCC-2 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the TLCC-2 protein to bind to or interact with a TLCC-2 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular Ca^{2+} , diacylglycerol, IP_3 , and the like), detecting catalytic/enzymatic activity of the target using an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response.

[0524] In yet another embodiment, an assay of the present invention is a cell-free assay in which a TLCC-2 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the TLCC-2 protein or biologically active portion thereof is determined. Preferred biologically active portions of the TLCC-2 proteins to be used in assays of the present invention include fragments which participate in interactions with non-TLCC-2 molecules, *e.g.*, fragments with high surface probability scores. Binding of the test compound to the TLCC-2 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the TLCC-2 protein or biologically active portion thereof with a known compound which binds TLCC-2 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a TLCC-2 protein, wherein determining the ability of the test compound to interact with a TLCC-2 protein comprises determining the ability of the test compound to preferentially bind to TLCC-2 or biologically active portion thereof as compared to the known compound.

[0525] In another embodiment, the assay is a cell-free assay in which a TLCC-2 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the TLCC-2 protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a TLCC-2 protein can be accomplished, for example, by determining the ability of the TLCC-2 protein to bind to a TLCC-2 target molecule by one of the methods described above for determining direct binding. Determining the ability of the TLCC-2 protein to bind to a TLCC-2 target molecule can also

be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

[0526] In an alternative embodiment, determining the ability of the test compound to modulate the activity of a TLCC-2 protein can be accomplished by determining the ability of the TLCC-2 protein to further modulate the activity of a downstream effector of a TLCC-2 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

[0527] In yet another embodiment, the cell-free assay involves contacting a TLCC-2 protein or biologically active portion thereof with a known compound which binds the TLCC-2 protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the TLCC-2 protein, wherein determining the ability of the test compound to interact with the TLCC-2 protein comprises determining the ability of the TLCC-2 protein to preferentially bind to or modulate the activity of a TLCC-2 target molecule.

[0528] In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either TLCC-2 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a TLCC-2 protein, or interaction of a TLCC-2 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ TLCC-2 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or TLCC-2 protein, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological

conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of TLCC-2 binding or activity determined using standard techniques.

[0529] Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a TLCC-2 protein or a TLCC-2 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated TLCC-2 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with TLCC-2 protein or target molecules but which do not interfere with binding of the TLCC-2 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or TLCC-2 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the TLCC-2 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the TLCC-2 protein or target molecule.

[0530] In another embodiment, modulators of TLCC-2 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of TLCC-2 mRNA or protein in the cell is determined. The level of expression of TLCC-2 mRNA or protein in the presence of the candidate compound is compared to the level of expression of TLCC-2 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of TLCC-2 expression based on this comparison. For example, when expression of TLCC-2 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of TLCC-2 mRNA or protein expression. Alternatively, when expression of TLCC-2 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of TLCC-2 mRNA or protein expression. The level of TLCC-2 mRNA or protein expression in the cells can be determined by methods described herein for detecting TLCC-2 mRNA or protein.

[0531] In yet another aspect of the invention, the TLCC-2 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with TLCC-2 ("TLCC-2-binding proteins" or "TLCC-2-bp") and are involved in TLCC-2 activity. Such TLCC-2-binding proteins are also likely to be involved in the propagation of signals by the TLCC-2 proteins or TLCC-2 targets as, for example, downstream elements of a TLCC-2-mediated signaling pathway. Alternatively, such TLCC-2-binding proteins are likely to be TLCC-2 inhibitors.

[0532] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a TLCC-2 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a TLCC-2-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the TLCC-2 protein.

[0533] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a TLCC-2 protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

[0534] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a TLCC-2 modulating agent, an antisense TLCC-2

nucleic acid molecule, a TLCC-2-specific antibody, or a TLCC-2-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

Detection Assays

[0535] Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[0536] Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the TLCC-2 nucleotide sequences, described herein, can be used to map the location of the TLCC-2 genes on a chromosome. The mapping of the TLCC-2 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

[0537] Briefly, TLCC-2 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the TLCC-2 nucleotide sequences. Computer analysis of the TLCC-2 sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the TLCC-2 sequences will yield an amplified fragment.

[0538] Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse

chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

[0539] PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the TLCC-2 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a TLCC-2 sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

[0540] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

[0541] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are

more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[0542] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature*, 325:783-787.

[0543] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the TLCC-2 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

[0544] The TLCC-2 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[0545] Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the TLCC-2 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences.

These primers can then be used to amplify an individual's DNA and subsequently sequence it.

[0546] Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The TLCC-2 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:4 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:6 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[0547] If a panel of reagents from TLCC-2 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of TLCC-2 Sequences in Forensic Biology

[0548] DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[0549] The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:4 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the TLCC-2 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:4 having a length of at least 20 bases, preferably at least 30 bases.

[0550] The TLCC-2 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such TLCC-2 probes can be used to identify tissue by species and/or by organ type.

[0551] In a similar fashion, these reagents, *e.g.*, TLCC-2 primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine:

[0552] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining TLCC-2 protein and/or nucleic acid expression as well as TLCC-2 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted TLCC-2 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with TLCC-2 protein, nucleic acid expression or

activity. For example, mutations in a TLCC-2 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with TLCC-2 protein, nucleic acid expression or activity.

[0553] Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of TLCC-2 in clinical trials.

[0554] These and other agents are described in further detail in the following sections.

Diagnostic Assays

[0555] An exemplary method for detecting the presence or absence of TLCC-2 protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting TLCC-2 protein or nucleic acid (*e.g.*, mRNA, or genomic DNA) that encodes TLCC-2 protein such that the presence of TLCC-2 protein or nucleic acid is detected in the biological sample. A preferred agent for detecting TLCC-2 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to TLCC-2 mRNA or genomic DNA. The nucleic acid probe can be, for example, the TLCC-2 nucleic acid set forth in SEQ ID NO:4 or 6, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to TLCC-2 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

[0556] A preferred agent for detecting TLCC-2 protein is an antibody capable of binding to TLCC-2 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells

and fluids present within a subject. That is, the detection method of the invention can be used to detect TLCC-2 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of TLCC-2 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of TLCC-2 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of TLCC-2 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of TLCC-2 protein include introducing into a subject a labeled anti-TLCC-2 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[0557] In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

[0558] In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting TLCC-2 protein, mRNA, or genomic DNA, such that the presence of TLCC-2 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of TLCC-2 protein, mRNA or genomic DNA in the control sample with the presence of TLCC-2 protein, mRNA or genomic DNA in the test sample.

[0559] The invention also encompasses kits for detecting the presence of TLCC-2 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting TLCC-2 protein or mRNA in a biological sample; means for determining the amount of TLCC-2 in the sample; and means for comparing the amount of TLCC-2 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect TLCC-2 protein or nucleic acid.

Prognostic Assays

[0560] The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted TLCC-2 expression or activity. As used herein, the term "aberrant" includes a TLCC-2 expression or activity which deviates from the wild type TLCC-2

expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant TLCC-2 expression or activity is intended to include the cases in which a mutation in the TLCC-2 gene causes the TLCC-2 gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional TLCC-2 protein or a protein which does not function in a wild-type fashion, *e.g.*, a protein which does not interact with a TLCC-2 substrate, *e.g.*, a non-calcium channel subunit or ligand, or one which interacts with a non-TLCC-2 substrate, *e.g.* a non-calcium channel subunit or ligand. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response, such as cellular proliferation. For example, the term unwanted includes a TLCC-2 expression or activity which is undesirable in a subject.

[0561] The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in TLCC-2 protein activity or nucleic acid expression, such as a CNS disorder (*e.g.*, a neurodegenerative disorder, a pain disorder, or a cellular proliferation, growth, differentiation, or migration disorder). Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in TLCC-2 protein activity or nucleic acid expression, such as a CNS disorder, a pain disorder, or a cellular proliferation, growth, differentiation, or migration disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted TLCC-2 expression or activity in which a test sample is obtained from a subject and TLCC-2 protein or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of TLCC-2 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted TLCC-2 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

[0562] Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted TLCC-2 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an

agent for a CNS disorder, pain disorder, or a cellular proliferation, growth, differentiation, or migration disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted TLCC-2 expression or activity in which a test sample is obtained and TLCC-2 protein or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of TLCC-2 protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted TLCC-2 expression or activity).

[0563] The methods of the invention can also be used to detect genetic alterations in a TLCC-2 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in TLCC-2 protein activity or nucleic acid expression, such as a CNS disorder, pain disorder, or a disorder of cellular growth, differentiation, or migration. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a TLCC-2 -protein, or the mis-expression of the TLCC-2 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a TLCC-2 gene; 2) an addition of one or more nucleotides to a TLCC-2 gene; 3) a substitution of one or more nucleotides of a TLCC-2 gene, 4) a chromosomal rearrangement of a TLCC-2 gene; 5) an alteration in the level of a messenger RNA transcript of a TLCC-2 gene, 6) aberrant modification of a TLCC-2 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a TLCC-2 gene, 8) a non-wild type level of a TLCC-2-protein, 9) allelic loss of a TLCC-2 gene, and 10) inappropriate post-translational modification of a TLCC-2-protein. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a TLCC-2 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

[0564] In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the TLCC-2-gene (see Abravaya *et al.* (1995) *Nucleic*

Acids Res. 23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a TLCC-2 gene under conditions such that hybridization and amplification of the TLCC-2-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

[0565] Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

[0566] In an alternative embodiment, mutations in a TLCC-2 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[0567] In other embodiments, genetic mutations in TLCC-2 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in TLCC-2 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays

of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[0568] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the TLCC-2 gene and detect mutations by comparing the sequence of the sample TLCC-2 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

[0569] Other methods for detecting mutations in the TLCC-2 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type TLCC-2 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

[0570] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called

“DNA mismatch repair” enzymes) in defined systems for detecting and mapping point mutations in TLCC-2 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a TLCC-2 sequence, *e.g.*, a wild-type TLCC-2 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

[0571] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in TLCC-2 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control TLCC-2 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

[0572] In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[0573] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective

primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

[0574] Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[0575] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a TLCC-2 gene.

[0576] Furthermore, any cell type or tissue in which TLCC-2 is expressed may be utilized in the prognostic assays described herein.

Monitoring of Effects During Clinical Trials

[0577] Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a TLCC-2 protein (*e.g.*, the modulation of membrane excitability) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase TLCC-2 gene expression,

protein levels, or upregulate TLCC-2 activity, can be monitored in clinical trials of subjects exhibiting decreased TLCC-2 gene expression, protein levels, or downregulated TLCC-2 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease TLCC-2 gene expression, protein levels, or downregulate TLCC-2 activity, can be monitored in clinical trials of subjects exhibiting increased TLCC-2 gene expression, protein levels, or upregulated TLCC-2 activity. In such clinical trials, the expression or activity of a TLCC-2 gene, and preferably, other genes that have been implicated in, for example, a TLCC-2-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

[0578] For example, and not by way of limitation, genes, including TLCC-2, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates TLCC-2 activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on TLCC-2-associated disorders (*e.g.*, disorders characterized by deregulated signaling or membrane excitation), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of TLCC-2 and other genes implicated in the TLCC-2-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of TLCC-2 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

[0579] In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a TLCC-2 protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the TLCC-2 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the TLCC-2 protein, mRNA, or genomic DNA in the pre-administration sample

with the TLCC-2 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of TLCC-2 to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of TLCC-2 to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, TLCC-2 expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

Methods of Treatment:

[0580] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted TLCC-2 expression or activity, *e.g.* a CNS disorder, pain disorder, or a cellular proliferation, growth, differentiation, or migration disorder. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the TLCC-2 molecules of the present invention or TLCC-2 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

Prophylactic Methods

[0581] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted TLCC-2 expression or activity, by administering to the subject a TLCC-2 or an agent which modulates TLCC-2 expression

or at least one TLCC-2 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted TLCC-2 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the TLCC-2 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of TLCC-2 aberrancy, for example, a TLCC-2, TLCC-2 agonist or TLCC-2 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Therapeutic Methods

[0582] Another aspect of the invention pertains to methods of modulating TLCC-2 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a TLCC-2 or agent that modulates one or more of the activities of TLCC-2 protein activity associated with the cell. An agent that modulates TLCC-2 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a TLCC-2 protein (*e.g.*, a TLCC-2 substrate), a TLCC-2 antibody, a TLCC-2 agonist or antagonist, a peptidomimetic of a TLCC-2 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more TLCC-2 activities. Examples of such stimulatory agents include active TLCC-2 protein and a nucleic acid molecule encoding TLCC-2 that has been introduced into the cell. In another embodiment, the agent inhibits one or more TLCC-2 activities. Examples of such inhibitory agents include antisense TLCC-2 nucleic acid molecules, anti-TLCC-2 antibodies, and TLCC-2 inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a TLCC-2 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) TLCC-2 expression or activity. In another embodiment, the method involves administering a TLCC-2 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted TLCC-2 expression or activity.

[0583] Stimulation of TLCC-2 activity is desirable in situations in which TLCC-2 is abnormally downregulated and/or in which increased TLCC-2 activity is likely to have a beneficial effect. Likewise, inhibition of TLCC-2 activity is desirable in situations in which TLCC-2 is abnormally upregulated and/or in which decreased TLCC-2 activity is likely to have a beneficial effect.

Pharmacogenomics

[0584] The TLCC-2 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on TLCC-2 activity (*e.g.*, TLCC-2 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) TLCC-2-associated disorders (*e.g.*, proliferative disorders) associated with aberrant or unwanted TLCC-2 activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a TLCC-2 molecule or TLCC-2 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a TLCC-2 molecule or TLCC-2 modulator.

[0585] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11): 983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[0586] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[0587] Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drugs target is known (*e.g.*, a TLCC-2 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[0588] As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor

metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

[0589] Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a TLCC-2 molecule or TLCC-2 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[0590] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a TLCC-2 molecule or TLCC-2 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[0591] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Sequence Listing, are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN TLCC-2 cDNA

[0592] In this example, the identification and characterization of the gene encoding human TLCC-2 (clone Fbh54420FL) is described.

Isolation of the TLCC-2 cDNA

[0593] The invention is based, at least in part, on the discovery of a human gene encoding a novel protein, referred to herein as TLCC-2. The entire sequence of the human clone Fbh54420FL was determined and found to contain an open reading frame termed human "TLCC-2". The nucleotide sequence of the human TLCC-2 gene is set forth in the Sequence Listing as SEQ ID NO:4 and 6. The amino acid sequence of the human TLCC-2 expression product is set forth in the Sequence Listing as SEQ ID NO:5.

[0594] The nucleotide sequence encoding the human TLCC-2 protein is set forth as SEQ ID NO:4. The protein encoded by this nucleic acid comprises about 580 amino acids and has the amino acid sequence is set forth as SEQ ID NO:5. The coding region (open reading frame) of SEQ ID NO:4 is set forth as SEQ ID NO:6.

Analysis of the Human TLCC-2 Molecules

[0595] A BLASTN 2.0 search against the dbEST database, using a score of 100 and a wordlength of 12 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403) of the nucleotide sequence of human TLCC-2 revealed that human IC54420 is 99% identical to tb24a12.x1 NCI_CGAP_kid12 *Homo sapiens* cDNA clone IMAGE:2055262 3' similar to WP:R13A5.1 CE01370 (Accession Number AI307240) over nucleotides 2077 to 1377. The search further revealed that human TLCC-2 is 98% identical to au44h03.x1 Schneider fetal brain 00004 *Homo sapiens* cDNA clone IMAGE:2517653 3' similar to WP:R13A5.1 CE01370 (Accession Number AI816064) over nucleotides 2079-1375. This search further revealed that human TLCC-2 is 97% identical to wv36f01.x1 NCI_CGAP_Ov18 *Homo sapiens* cDNA clone IMAGE:2531641 3' similar to WP:R13A5.1 CE01370 (Accession Number AI951554) over nucleotides 2088 to 1407. This search further revealed that human TLCC-2 is 97% identical to wp80f10.x1 NCI_CGAP_Brn25 *Homo sapiens* cDNA clone IMAGE:2468107 3' similar to WP:R13A5.1 CE01370 (Accession Number AI942492) over nucleotides 2072-1422. The search further revealed that human TLCC-2 is 96% identical to nr72c11.s1 NCI_CGAP_Pr24 *Homo sapiens* cDNA clone IMAGE:1173524 similar to WP:R13A5.1 CE01370 (Accession Number AA641031) over nucleotides 2073-1407.

[0596] A BLASTX 2.0 search against the NRP/protot database, using a score of 100 and a wordlength of 3 (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403), of the translated nucleotide sequence of human TLCC-2 revealed that human TLCC-2 is 67% or less identical to fragments (*e.g.*, fragments of 185 amino acids or less) of unnamed protein

product from *Homo sapiens* (Accession Number AK001868). The search further revealed that human TLCC-2 is 27% or less identical to fragments (*e.g.*, fragments of 75 amino acids or less) of human polycystic kidney disease and receptor for egg jelly related protein (Accession Number AF116458). This search further revealed that human TLCC-2 is 43% or less identical to fragments (*e.g.*, fragments of 59 amino acids or less) of polycystic kidney disease and receptor for egg jelly related protein from *Mus musculus* (Accession Number AF116459). This search further revealed that human TLCC-2 is 70% or less identical to fragments (*e.g.*, fragments of 29 amino acids or less) of a 110 amino acid long hypothetical protein from *Aeropyrum pernix* (Accession Number AP000060) over the full length of this protein.

[0597] An alignment of the human TLCC-2 amino acid sequence with the amino acid sequences of vanilloid receptor 1 and vanilloid receptor 2 from *Homo sapiens* (Accession Numbers AJ277028 and AF103906, set forth as SEQ ID NO:9 and SEQ ID NO:10, respectively) using the CLUSTAL W (1.74) multiple sequence alignment program can be performed. An alignment of the human TLCC-2 amino acid sequence with the amino acid sequence of polycystic kidney disease protein 2 from *Mus musculus* (Accession Number NP032887, set forth as SEQ ID NO:11) using the CLUSTAL W(1.74) multiple sequence alignment program can be performed. An alignment of the human TLCC-2 amino acid sequence with the amino acid sequence of human melastatin (Accession Number AF071787, set forth as SEQ ID NO:12) using the CLUSTAL W(1.74) multiple sequence alignment program can be performed.

[0598] A search was performed against the Memsat database, resulting in the identification of six transmembrane domains in the amino acid sequence of human TLCC-2 (SEQ ID NO:5) at about residues 70-86, 299-317, 354-371, 385-416, 428-447, and 497-521.

[0599] A search was also performed against the Prosite database resulting in the identification of four N-glycosylation sites in the amino acid sequence of human TLCC-2 (SEQ ID NO:5) at about residues 159-162, 179-182, 220-223, and 230-233.

[0600] A search was also performed against the HMM database resulting in the identification of a fibronectin type III domain in the amino acid sequence of human TLCC-2 (SEQ ID NO:5) at about residues 202-269 (score = 5).

[0601] A search was also performed against the ProDom database resulting in the identification of a 31K RNA-4 protein domain (SEQ ID NO:8) in the amino acid sequence of human TLCC-2 (SEQ ID NO:5) at about residues 397-443 (score = 73).

Tissue Distribution of TLCC-2 mRNA by PCR

[0602] This example describes the tissue distribution of TLCC-2 mRNA, as was determined by Polymerase Chain Reaction (PCR) on cDNA libraries using oligonucleotide primers based on the human TLCC-2 sequence. The human TLCC-2 gene was found to be expressed in adrenal gland, in blood, in brain, in breast, in d8 dendritic cells, in endothelial cells, in all fetal tissues, in heart, in keratinocytes, in lymphocytes, in muscle, in placenta, in prostate, in spinal cord, in spleen, in T cells, and in testis.

Tissue Distribution of Human TLCC-2 mRNA by Northern Analysis

[0603] This example describes the tissue distribution of TLCC-2 mRNA, as determined by Northern analysis.

[0604] Northern blot hybridizations with the various RNA samples are performed under standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65°C. The DNA probe is radioactively labeled with 32P-dCTP using the Prime-It kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing human mRNA (MultiTissue Northern I and MultiTissue Northern II from Clontech, Palo Alto, CA) are probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Tissue Distribution of TLCC-2 mRNA by In situ Analysis

[0605] For in situ analysis, various tissues, e.g. tissues obtained from brain and spinal cord from monkey and rat, were first frozen on dry ice. Ten-micrometer-thick sections of the tissues were post-fixed with 4% formaldehyde in DEPC treated 1X phosphate-buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections were rinsed in DEPC 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissue was then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

[0606] Hybridizations were performed with 35S-radiolabeled (5 X 10⁷ cpm/ml) cRNA probes. Probes were incubated in the presence of a solution containing 600 mM

NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

[0607] After hybridization, slides were washed with 2X SSC. Sections are then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10µg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides were then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections were then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

[0608] In situ hybridization results showed expression in monkey and rat brain (including cortex, stratum, and hippocampus), spinal cord, and dorsal root ganglia (DRG) neurons.

[0609] In situ hybridization in rat animal models showed up-regulation of the TLCC-2 gene 10 days after unilateral chronic constriction injury (CCI). There was up-regulation of TLCC-2 seven days after axotomy and after intraplantar injection of complete Freund's adjuvant (CFA). These levels decreased to normal levels at later time points. No contralateral effects were observed. These results indicate that the TLCC-2 molecules of the present invention are up-regulated in response to painful stimuli, and are therefore involved in nociception. Modulation, e.g., inhibition, of expression or activity of the TLCC-2 molecules of the invention may therefore modulate nociception and provide treatment for pain disorders.

Tissue Expression Analysis of TLCC-2 mRNA Using Taqman Analysis

[0610] This example describes the tissue distribution of human TLCC-2 mRNA in a variety of cells and tissues, as determined using the TaqMan™ procedure. The Taqman™ procedure is a quantitative, reverse transcription PCR-based approach for detecting mRNA. The RT-PCR reaction exploits the 5' nuclease activity of AmpliTaq Gold™ DNA Polymerase to cleave a TaqMan™ probe during PCR. Briefly, cDNA was generated from the samples of interest, e.g., human brain, spinal cord, heart, kidney, liver, lung, dorsal root

ganglia, and skin, and used as the starting material for PCR amplification. In addition to the 5' and 3' gene-specific primers, a gene-specific oligonucleotide probe (complementary to the region being amplified) was included in the reaction (i.e., the Taqman™ probe). The TaqMan™ probe includes the oligonucleotide with a fluorescent reporter dye covalently linked to the 5' end of the probe (such as FAM (6-carboxyfluorescein), TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein), JOE (6-carboxy-4,5-dichloro-2,7-dimethoxyfluorescein), or VIC) and a quencher dye (TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) at the 3' end of the probe.

[0611] During the PCR reaction, cleavage of the probe separates the reporter dye and the quencher dye, resulting in increased fluorescence of the reporter. Accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe is intact, the proximity of the reporter dye to the quencher dye results in suppression of the reporter fluorescence. During PCR, if the target of interest is present, the probe specifically anneals between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the AmpliTaq™ Gold DNA Polymerase cleaves the probe between the reporter and the quencher only if the probe hybridizes to the target. The probe fragments are then displaced from the target, and polymerization of the strand continues. The 3' end of the probe is blocked to prevent extension of the probe during PCR. This process occurs in every cycle and does not interfere with the exponential accumulation of product. RNA was prepared using the trizol method and treated with DNase to remove contaminating genomic DNA. cDNA was synthesized using standard techniques. Mock cDNA synthesis in the absence of reverse transcriptase resulted in samples with no detectable PCR amplification of the control gene confirms efficient removal of genomic DNA contamination.

[0612] Two human normal tissue panels indicated broad distribution of human TLCC-2 expression, with highest expression in human brain, followed by testis, placenta, adrenal gland, spinal cord, skin, and dorsal root ganglia (DRG).

EXAMPLE 2: EXPRESSION OF RECOMBINANT IC54420 PROTEIN IN BACTERIAL CELLS

[0613] In this example, TLCC-2 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, TLCC-2 is fused to GST and this fusion polypeptide is

expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-TLCC-2 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 3: EXPRESSION OF RECOMBINANT IC54420 PROTEIN IN COS CELLS

[0614] To express the TLCC-2 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire TLCC-2 protein and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

[0615] To construct the plasmid, the TLCC-2 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the TLCC-2 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the TLCC-2 coding sequence. The PCR amplified fragment and the pcDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the TLCC-2 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

[0616] COS cells are subsequently transfected with the TLCC-2-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods,

DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the IC54420 polypeptide is detected by radiolabelling (^{35}S -methionine or ^{35}S -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S -methionine (or ^{35}S -cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

[0617] Alternatively, DNA containing the TLCC-2 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the TLCC-2 polypeptide is detected by radiolabelling and immunoprecipitation using a TLCC-2-specific monoclonal antibody.

EXAMPLE 4: REGULATION OF CALCIUM INFLUX THROUGH TLCC-2

[0618] This experiment describes the regulation of calcium influx though TLCC-2 in HEK293 cells as determined by Fluorometric Imaging Plate Reader (FLIPR) (Molecular Devices Corp., Sunnyvale, CA) experiments.

[0619] The FLIPR is a screening tool for cell-based fluorescent assays which allows the simultaneous stimulation and measurement of separate cell populations in a high throughput format. Therefore, it is possible to quantify transient signals, such as the release of intracellular calcium, from cell populations, in parallel and in real time. The FLIPR contains chambers in which to hold the test plate and plates containing antagonists or agonists to be added to the test plate. The FLIPR utilizes an argon laser that provides discrete spectral lines spaced from approximately 350 to 530 nm. For use with fluorescent Ca^{2+} dyes, the 88-nm line of the laser is employed. The laser simultaneously illuminates the wells in a test plate. The image of each well in the plate is captured by a cooled charge

coupled device (CCD) camera, which updates images once per second, if required, for the measurement of rapid calcium responses. Because both excitation and emission are read via the bottom of the plate, black-walled, transparent bottomed 96-well plates are used. Data captured by the CCD camera is converted to digital data and then transferred to a computer.

[0620] Briefly, a calcium indicator (e.g., fluo-3/AM or Calcium Green-1/AM) was transferred to the culture medium. Because the FLIPR collects fluorescence from the bottom of the well, suspension cells require centrifugation to the base of the well following dye loading. Viable HEK293 cells were resuspended in loading medium and incubated for one hour. The cells were then centrifuged and resuspended with wash buffer. The cell suspension containing the dye was then aliquotted into each well of the black-walled, transparent bottomed 96-well plate and the plate was centrifuged. Adherent cells may be plated at an appropriate density in the 96-well plates and cultured overnight. Dye is then loaded and incubated. The FLIPR assay was then carried out and the results analyzed.

[0621] Results show a constitutive calcium influx through TLCC-2 in HEK293 cells that were incubated with NMDG/0 Ca²⁺ and stimulated afterwards with 5mM Ca²⁺.

Equivalents

[0622] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

III. 8797, A NOVEL HUMAN GALACTOSYLTRANSFERASE AND USES THEREOF

Background of the Invention

[0623] Glycoproteins are glycosylated by the cotranslational addition of carbohydrates (*i.e.*, sugars) to specific amino acid residues on the protein (Imperiali *et al.* (1999) *Curr. Opin. Cell Biol.* 3:643-649). After transfer, the sugars are processed by the actions of glycosylhydrolases, which trim sugars, and glycosyltransferases, which add sugars. The glycosylation of proteins can dramatically alter the folding (*i.e.*, the structure)

and, therefore, the function of the protein. This modification also serves to stabilize the protein, as well as to assist in the assembly of oligomeric complexes and the correct orientation of cell surface glycoproteins at the plasma membrane.

[0624] Glycosyltransferases are a family of enzymes which catalyze the formation of a glycosidic bond between two sugar molecules (*e.g.*, a nucleotide-bound donor sugar and an acceptor-bound acceptor sugar) (Darnell *et al.*, *Molecular Cell Biology*, Scientific American Books, Inc., 1990; Voet and Voet, *Biochemistry*, John Wiley and Sons, Inc., 1990). These enzymes have a precise specificity for substrate, donor sugar nucleotide, and acceptor. Members of this family of enzymes vary in structure, although glycosyltransferases share several characteristics. Glycosyltransferases are integral membrane proteins that possess a short amino-terminal cytoplasmic domain, a transmembrane domain, and a larger carboxy-terminal catalytic domain that typically consists of 325 or more amino acids (Natsuka *et al.* (1994) *Curr. Opin. Struct. Biol.* 4:683-691). Although most of these proteins are membrane bound, they may be proteolytically cleaved into soluble forms which may be secreted.

[0625] Glycosyltransferase sugar specificity may be directed to sugars such as galactose, glucose, fucose, or mannose, by galactosyltransferases, glucosyltransferases, fucosyltransferases, or mannosyltransferases, respectively (for a review, see the WWW Guide to Cloned Glycosyltransferases, available online through Wilson, I., Institut für Chemie der Universität für Bodenkultur, Muthgasse 18, Wein (1996)). Galactosyltransferases are involved in lactose synthesis and transfer galactose to N-acetylglucosamine, yielding N-acetyllactosamine (Voet and Voet, *Biochemistry*, John Wiley and Sons, Inc., 1990). The transfer of galactose may be directed to a growing oligosaccharide, lipid, or protein acceptor (Breton *et al.* (1999) *Curr. Opin. Struct. Biol.* 9:563-571). These enzymes are typically found in the trans Golgi, although they may occasionally be located to the cell surface or in soluble forms in milk, amniotic fluid, cerebrospinal fluid, saliva, urine, and serum (Axford (1999) *Biochim. Biophys. Acta* 1455:219-229). Galactosyltransferases play a multifunctional role in normal cell physiology. They are expressed in a tissue specific manner, and are regulated in healthy tissues as well as in disease states. These enzymes are present on the cell surface of sperm, and play a role in mammary gland morphogenesis and lactation (Brockhausen *et al.* (1998) *Acta Anatomica* 161:36-78).

Summary of the Invention

[0626] The present invention is based, at least in part, on the discovery of novel human galactosyltransferase family members, referred to herein as “human galactosyltransferase-1” or “HGT-1” nucleic acid and polypeptide molecules. The HGT-1 nucleic acid and polypeptide molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, *e.g.*, cell physiology, and/or cellular proliferation, growth, differentiation, and/or migration. The present invention is also based, at least in part, on the discovery that the HGT-1 molecules of the present invention are differentially expressed (*e.g.*, upregulated) in different types of tumor cells, *e.g.*, breast, lung, and colon tumor cells. The present invention is still further based, at least in part, on the discovery that the HGT-1 molecules of the present invention are upregulated during the progression from attachment-dependent to attachment-independent growth of pre-malignant and malignant cells (*e.g.*, breast cells).

[0627] Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding HGT-1 polypeptides or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of HGT-1-encoding nucleic acids.

[0628] In one embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence set forth in SEQ ID NO:13 or SEQ ID NO:15. In another embodiment, the invention features an isolated nucleic acid molecule that encodes a polypeptide including the amino acid sequence set forth in SEQ ID NO:14.

[0629] In still other embodiments, the invention features isolated nucleic acid molecules including nucleotide sequences that are substantially identical (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical) to the nucleotide sequence set forth as SEQ ID NO:13 or SEQ ID NO:15. The invention further features isolated nucleic acid molecules including at least 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 615, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 3950, or 4000 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO:13 or SEQ

ID NO:15. In another embodiment, the invention features isolated nucleic acid molecules which encode a polypeptide including an amino acid sequence that is substantially identical (*e.g.*, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical) to the amino acid sequence set forth as SEQ ID NO:14. The present invention also features nucleic acid molecules which encode allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:14. In addition to isolated nucleic acid molecules encoding full-length polypeptides, the present invention also features nucleic acid molecules which encode fragments, for example, biologically active or antigenic fragments, of the full-length polypeptides of the present invention (*e.g.*, fragments including at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, or 375 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:14). In still other embodiments, the invention features nucleic acid molecules that are complementary to, antisense to, or hybridize under stringent conditions to the isolated nucleic acid molecules described herein.

[0630] In another aspect, the invention provides vectors including the isolated nucleic acid molecules described herein (*e.g.*, HGT-1-encoding nucleic acid molecules). Such vectors can optionally include nucleotide sequences encoding heterologous polypeptides. Also featured are host cells including such vectors (*e.g.*, host cells including vectors suitable for producing HGT-1 nucleic acid molecules and polypeptides).

[0631] In another aspect, the invention features isolated HGT-1 polypeptides and/or biologically active or antigenic fragments thereof. Exemplary embodiments feature a polypeptide including the amino acid sequence set forth as SEQ ID NO:14, a polypeptide including an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the amino acid sequence set forth as SEQ ID NO:14, a polypeptide encoded by a nucleic acid molecule including a nucleotide sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identical to the nucleotide sequence set forth as SEQ ID NO:13 or SEQ ID NO:15. Also featured are fragments of the full-length polypeptides described herein (*e.g.*, fragments including at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, or 375 contiguous amino acid residues of the sequence set forth as SEQ ID NO:14) as well

as allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:14.

[0632] The HGT-1 polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of galactosyltransferase associated disorders and/or cellular proliferation, growth, differentiation, and/or migration disorders. In one embodiment, an HGT-1 polypeptide or fragment thereof, has an HGT-1 activity. In another embodiment, an HGT-1 polypeptide or fragment thereof, has a transmembrane domain and/or a galactosyltransferase family domain, and optionally, has an HGT-1 activity. In a related aspect, the invention features antibodies (*e.g.*, antibodies which specifically bind to any one of the polypeptides described herein) as well as fusion polypeptides including all or a fragment of a polypeptide described herein.

[0633] The present invention further features methods for detecting HGT-1 polypeptides and/or HGT-1 nucleic acid molecules, such methods featuring, for example, a probe, primer or antibody described herein. Also featured are kits *e.g.*, kits for the detection of HGT-1 polypeptides and/or HGT-1 nucleic acid molecules. In a related aspect, the invention features methods for identifying compounds which bind to and/or modulate the activity of an HGT-1 polypeptide or HGT-1 nucleic acid molecule described herein. Further featured are methods for modulating an HGT-1 activity.

[0634] In other embodiments, the invention provides methods for identifying a subject having a cellular proliferation, growth, differentiation, and/or migration disorder, or at risk for developing a cellular proliferation, growth, differentiation, and/or migration disorder; methods for identifying a compound capable of treating a cellular proliferation, growth, differentiation, and/or migration disorder characterized by aberrant HGT-1 nucleic acid expression or HGT-1 polypeptide activity; and methods for treating a subject having a cellular proliferation, growth, differentiation, and/or migration disorder characterized by aberrant HGT-1 polypeptide activity or aberrant HGT-1 nucleic acid expression

[0635] Other features and advantages of the invention will be apparent from the following detailed description and claims.

Detailed Description of the Invention

[0636] The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as “human galactosyltransferase-1” or “HGT-1” nucleic acid and polypeptide molecules, which are novel members of the galactosyltransferase family. These novel molecules are capable of forming a glycosidic bond between molecules, *e.g.*, between UDP-galactose and N-acetylglucosamine and, thus, play a role in or function in a variety of cellular processes, *e.g.*, maintenance of cell physiology and lactose homeostasis, and/or cellular proliferation, growth, differentiation, and/or migration. The present invention is also based, at least in part, on the discovery that the HGT-1 molecules of the present invention are differentially expressed (*e.g.*, upregulated) in different types of tumor cells, *e.g.*, breast, lung, and colon tumor cells. the present invention is still further based, at least in part, on the discovery that the HGT-1 molecules of the present invention are upregulated during the progression from attachment-dependent to attachment-independent growth of pre-malignant and malignant cells (*e.g.*, breast cells).

[0637] As used herein, a “galactosyltransferase” includes a protein or polypeptide which is involved in forming a glycosidic bond between molecules, *e.g.*, between UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine on a polysaccharide or glycoprotein), in a cell (*e.g.*, in the Golgi complex (*e.g.*, the *trans* Golgi)).

Galactosyltransferase family members regulate lactose homeostasis in a cell (*i.e.*, via the formation of a glycosidic bond between galactose and glucose molecules) and, typically, have UDP-galactose specificity. Galactosyltransferase family members share a common topology: they are integral membrane proteins that possess a short amino-terminal cytoplasmic domain, a transmembrane domain, a stem region of variable length, and a carboxy-terminal catalytic domain. Although most members of this family are membrane bound, they may be proteolytically cleaved into soluble forms which may be secreted.

[0638] As used herein, a “galactosyltransferase mediated activity” includes an activity which involves a galactosyltransferase in a cell (*e.g.*, in the Golgi complex (*e.g.*, the *trans* Golgi)). Galactosyltransferase mediated activities include formation of a glycosidic bond between molecules, *e.g.*, between UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine on a polysaccharide or glycoprotein); regulation of lactose homeostasis; the participation in signal transduction pathways associated with oligosaccharide metabolism and glycoprotein glycosylation; and/or regulation of cellular differentiation, growth, differentiation, and/or migration.

[0639] As the HGT-1 molecules of the present invention are galactosyltransferases, they may be useful for developing novel diagnostic and therapeutic agents for galactosyltransferase associated disorders. As used herein, the term “galactosyltransferase associated disorder” includes a disorder, disease, or condition which is characterized by an aberrant, *e.g.*, upregulated or downregulated, galactosyltransferase mediated activity. Galactosyltransferase associated disorders typically result in upregulated or downregulated, oligosaccharide levels in a cell. Examples of galactosyltransferase associated disorders include disorders associated with oligosaccharide homeostasis, such as rheumatoid arthritis, juvenile chronic arthritis, Sjorgren’s syndrome, permanent mixed-field polyagglutinability, leukemia, lymphoma, colon cancer, and breast cancer.

[0640] As demonstrated herein, the HGT-1 molecules of the present invention are differentially expressed (*e.g.*, upregulated) in different types of tumor cells. Accordingly, the HGT-1 molecules of the present invention may be useful for developing novel diagnostic and therapeutic agents for cellular proliferation, growth, differentiation, and/or migration disorders. As used herein, “cellular proliferation, growth, differentiation, and/or migration disorders” include those disorders that affect cellular proliferation, growth, differentiation, and/or migration processes. As used herein, a “cellular proliferation, growth, differentiation, and/or migration process” is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. Examples of cellular proliferation, growth, differentiation, and/or migration disorders include cancer, *e.g.*, ovarian cancer, breast cancer, colon cancer, lung cancer, brain cancer, as well as other types of carcinomas, sarcomas, lymphomas, and/or leukemias; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or myeloproliferative disorders.

[0641] As further demonstrated herein, the HGT-1 molecules of the present invention are differentially expressed (*e.g.*, upregulated or downregulated) in human umbilical vein endothelial cells (HUVECs) under conditions of shear stress, and in the heart of subjects and animal models suffering from congestive heart failure. As used herein, the term “cardiovascular disorder” includes a disorder, disease or condition which affects the cardiovascular system, *e.g.*, the heart or blood vessels. Cardiovascular disorders can detrimentally affect cellular functions such as calcium transport and inter- or intra-cellular communication; and tissue functions such as angiogenesis, vascular smooth muscle tone,

vascular function, and cardiac function. Examples of cardiovascular disorders include cardiovascular disorders include hypertension, arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, arrhythmia, atherosclerosis, transplant atherosclerosis, varicose veins, migraine headaches, cluster headaches, vascular disease, diabetic vascular disease, pulmonary vascular disease, peripheral vascular disease, renovascular hypertension, intravascular tumor, pulmonary vasculitis, vascular tone disorders in pregnancy, pulmonary capillaritis, peripheral arterial disease, idiopathic hypereosinophilic syndrome, aortic aneurysm, respiratory disease, vasospasm, systemic sclerosis, preeclampsia, graft vessel disease, cardiac allograft vasculopathy, vascular ischemic injury, familial amyloidotic polyneuropathy, acute atherosclerosis, cardiovascular disease, Kawasaki disease, ischemic syndromes, chronic heart failure, and fibrosis.

[0642] The term “family” when referring to the polypeptide and nucleic acid molecules of the invention is intended to mean two or more polypeptides or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first polypeptide of human origin, as well as other, distinct polypeptides of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, mouse or monkey polypeptides. Members of a family may also have common functional characteristics.

[0643] For example, the family of HGT-1 polypeptides comprise at least one “transmembrane domain.” As used herein, the term “transmembrane domain” includes an amino acid sequence of about 15-45 amino acid residues in length which spans the plasma membrane. More preferably, a transmembrane domain includes about at least 15, 20, 25, 30, 35, 40, or 45 amino acid residues and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an alpha-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids

of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, alanines, valines, phenylalanines, prolines or methionines. Transmembrane domains are described in, for example, Zagotta W.N. *et al.* (1996) *Annu. Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference. A MEMSAT analysis resulted in the identification of one transmembrane domain in the amino acid sequence of human HGT-1 (SEQ ID NO:14) at about residues 15-32.

[0644] Accordingly, HGT-1 polypeptides having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a transmembrane domain of human HGT-1 are within the scope of the invention.

[0645] In another embodiment, the Peptide Structure program in the GCG software package can be used to find structural, hydrophobic, and antigenic regions in the human HGT-1 sequence (SEQ ID NO:14).

[0646] In another embodiment, an HGT-1 molecule of the present invention is identified based on the presence of at least one “galactosyltransferase family domain.” As used herein, the term “galactosyltransferase family domain” includes a protein domain having at least about 100-300 amino acid residues, having a bit score of at least 100 when compared against a galactosyltransferase family domain Hidden Markov Model (HMM), and a galactosyltransferase mediated activity. Preferably, a galactosyltransferase family domain includes a polypeptide having an amino acid sequence of about 125-275, 150-250, 175-225, or more preferably, about 219 amino acid residues, a bit score of at least 140, 150, 160, or more preferably about 173.8, and a galactosyltransferase mediated activity. To identify the presence of a galactosyltransferase family domain in an HGT-1 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the PFAM HMM database). A PFAM galactosyltransferase family domain has been assigned the PFAM Accession PF01762. A search was performed against the PFAM HMM database resulting in the identification of a galactosyltransferase family domain in the amino acid sequence of human HGT-1 (SEQ ID NO:14) at about residues 102-321 of SEQ ID NO:14.

[0647] A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Methods Enzymol.* 183:146-159; Gribskov *et al.* (1987) *Proc. Natl.*

Acad. Sci. USA 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

[0648] Preferably a “galactosyltransferase family domain” has a “galactosyltransferase mediated activity” as described herein. For example, a galactosyltransferase family domain may have the ability to form a glycosidic bond between molecules, *e.g.*, between UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine on a polysaccharide or glycoprotein), in a cell (*e.g.*, in the Golgi complex (*e.g.*, the *trans* Golgi)); and the ability to regulate lactose homeostasis in a cell. Accordingly, identifying the presence of a “galactosyltransferase family domain” can include isolating a fragment of an HGT-1 molecule (*e.g.*, an HGT-1 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned galactosyltransferase mediated activities.

[0649] In a preferred embodiment, the HGT-1 molecules of the invention include at least one transmembrane domain and/or at least one galactosyltransferase family domain.

[0650] Isolated polypeptides of the present invention, preferably HGT-1 polypeptides, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:14 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:13 or 15. As used herein, the term “sufficiently identical” refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homology or identity and share a common functional activity are defined herein as sufficiently identical.

[0651] In a preferred embodiment, an HGT-1 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a galactosyltransferase family domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more homologous or identical to the amino acid sequence of SEQ ID NO:14. In yet another preferred embodiment, an HGT-1 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a galactosyltransferase family domain, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:13 or SEQ ID NO:15. In another preferred embodiment, an HGT-1 polypeptide includes at least one or more of the following domains: a transmembrane domain and/or a galactosyltransferase family domain, and has an HGT-1 activity.

[0652] As used interchangeably herein, an “HGT-1 activity”, “biological activity of HGT-1” or “functional activity of HGT-1”, includes an activity exerted by an HGT-1 polypeptide or nucleic acid molecule, for example, in an HGT-1 expressing cell or tissue, or on an HGT-1 target or substrate (*e.g.*, UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine bound to a polysaccharide and/or a glycoprotein)), as determined *in vivo* or *in vitro*, according to standard techniques. In one embodiment, an HGT-1 activity is a direct activity, such as association with or enzymatic modification of an HGT-1-target molecule. As used herein, a “target molecule” or “binding partner” is a molecule with which an HGT-1 polypeptide binds or interacts in nature, such that HGT-1-mediated function is achieved. An HGT-1 target molecule can be a non- HGT-1 molecule or an HGT-1 polypeptide of the present invention. In an exemplary embodiment, an HGT-1 target molecule is an HGT-1 substrate (*e.g.*, an UDP-galactose and N-acetylglucosamine). Furthermore, an HGT-1 activity can be an indirect activity, such as a cellular signaling activity mediated by interaction of the HGT-1 polypeptide with an HGT-1 substrate or binding partner. The biological activities of HGT-1 are described herein.

[0653] For example, an HGT-1 molecule can have one or more of the following activities: (i) it may bind UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine bound to a glycoprotein); (ii) it may catalyze the formation of glycosidic bonds (*e.g.*, between UDP-galactose and N-acetylglucosamine); (iii) it may modulate lactose homeostasis; (iv) it may regulate embryogenesis; (v) it may regulate development; (vi) it

may regulate the formation of structural elements of the cell; (vii) it may regulate the metabolism of adhesive ligands; (viii) it may regulate the metabolism of glycoprotein ligands and receptors; (ix) it may regulate blood clotting; (x) it may regulate thrombus dissolution; (xi) it may regulate hormone action; (xii) it may regulate fertilization; (xiii) it may regulate an immune system response; and (xiv) it may regulate cellular proliferation, growth, differentiation, and/or migration.

[0654] The nucleotide sequence of the isolated human HGT-1 cDNA and the predicted amino acid sequence of the human HGT-1 polypeptide are shown in SEQ ID NOs:13 and 14, respectively.

[0655] The human HGT-1 gene, which is approximately 4052 nucleotides in length, encodes a polypeptide having a molecular weight of approximately 41.6 kD and which is approximately 378 amino acid residues in length.

[0656] Various aspects of the invention are described in further detail in the following subsections:

Isolated Nucleic Acid Molecules

[0657] One aspect of the invention pertains to isolated nucleic acid molecules that encode HGT-1 polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify HGT-1-encoding nucleic acid molecules (*e.g.*, HGT-1 mRNA) and fragments for use as PCR primers for the amplification or mutation of HGT-1 nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0658] The term “isolated nucleic acid molecule” includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term “isolated” includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an “isolated” nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid

is derived. For example, in various embodiments, the isolated HGT-1 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0659] A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:13 or 15, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:13 or 15 as a hybridization probe, HGT-1 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

[0660] Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:13 or 15 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:13 or 15.

[0661] A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to HGT-1 nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

[0662] In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:13. The sequence of SEQ ID NO:13 corresponds to the human HGT-1 cDNA. This cDNA comprises sequences encoding the human HGT-1 polypeptide (*i.e.*, “the coding region”, from nucleotides 459-1592) as well as 5’ untranslated sequences (nucleotides 1-458) and 3’ untranslated sequences (nucleotides 1593-4052). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:13 (*e.g.*, nucleotides 459-1592, corresponding to SEQ ID NO:15). Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention comprises SEQ ID NO:15 and nucleotides 1-458 of SEQ ID NO:13. In yet

another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:15 and nucleotides 1593-4052 of SEQ ID NO:13. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:13 or SEQ ID NO:15. In still another embodiment, the nucleic acid molecule can comprise the coding region of SEQ ID NO:13 (*e.g.*, nucleotides 459-1592, corresponding to SEQ ID NO:15), as well as a stop codon (*e.g.*, nucleotides 1593-1595 of SEQ ID NO:13). In other embodiments, the nucleic acid molecule can comprise nucleotides 1-227, 658-748, 1142-1494, or 2149-2489 of SEQ ID NO:13.

[0663] In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:13 or 15, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:13 or 15 is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:13 or 15 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:13 or 15, thereby forming a stable duplex.

[0664] In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to the nucleotide sequence shown in SEQ ID NO:13 or 15 (*e.g.*, to the entire length of the nucleotide sequence), or a portion of any of these nucleotide sequences. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least (or no greater than) 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 615, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 3950, 4000 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:13 or 15.

[0665] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:13 or 15, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of an HGT-1 polypeptide,

e.g., a biologically active portion of an HGT-1 polypeptide. The nucleotide sequence determined from the cloning of the HGT-1 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other HGT-1 family members, as well as HGT-1 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The probe/primer (*e.g.*, oligonucleotide) typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, or 100 or more consecutive nucleotides of a sense sequence of SEQ ID NO:13 or 15, of an anti-sense sequence of SEQ ID NO:13 or 15, or of a naturally occurring allelic variant or mutant of SEQ ID NO:13 or 15.

[0666] Exemplary probes or primers are at least 12, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 or more nucleotides in length and/or comprise consecutive nucleotides of an isolated nucleic acid molecule described herein. Probes based on the HGT-1 nucleotide sequences can be used to detect (*e.g.*, specifically detect) transcripts or genomic sequences encoding the same or homologous polypeptides. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of an HGT-1 sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more nucleotides in length. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress an HGT-1 polypeptide, such as by measuring a level of an HGT-1-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting HGT-1 mRNA levels or determining whether a genomic HGT-1 gene has been mutated or deleted.

[0667] A nucleic acid fragment encoding a “biologically active portion of an HGT-1 polypeptide” can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:13 or 15, which encodes a polypeptide having an HGT-1 biological activity (the biological activities of the HGT-1 polypeptides are described herein), expressing the encoded portion of the HGT-1 polypeptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the HGT-1 polypeptide. In an exemplary embodiment, the nucleic acid molecule is at least 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 615, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200,

1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 3950, 4000 or more nucleotides in length and encodes a polypeptide having an HGT-1 activity (as described herein).

[0668] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:13 or 15. Such differences can be due to degeneracy of the genetic code, thus resulting in a nucleic acid which encodes the same HGT-1 polypeptides as those encoded by the nucleotide sequence shown in SEQ ID NO:13 or 15. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a polypeptide having an amino acid sequence which differs by at least 1, but no greater than 5, 10, 20, 50 or 100 amino acid residues from the amino acid sequence shown in SEQ ID NO:14. In yet another embodiment, the nucleic acid molecule encodes the amino acid sequence of human HGT-1. If an alignment is needed for this comparison, the sequences should be aligned for maximum homology.

[0669] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologues (different locus), and orthologues (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

[0670] Allelic variants result, for example, from DNA sequence polymorphisms within a population (*e.g.*, the human population) that lead to changes in the amino acid sequences of the HGT-1 polypeptides. Such genetic polymorphism in the HGT-1 genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules which include an open reading frame encoding an HGT-1 polypeptide, preferably a mammalian HGT-1 polypeptide, and can further include non-coding regulatory sequences, and introns.

[0671] Accordingly, in one embodiment, the invention features isolated nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:14, wherein the nucleic acid molecule hybridizes to a

complement of a nucleic acid molecule comprising SEQ ID NO:13 or SEQ ID NO:15, for example, under stringent hybridization conditions.

[0672] Allelic variants of human HGT-1 include both functional and non-functional HGT-1 polypeptides. Functional allelic variants are naturally occurring amino acid sequence variants of the human HGT-1 polypeptide that have an HGT-1 activity, *e.g.*, maintain the ability to bind an HGT-1 ligand or substrate and/or modulate galactosyltransferase activity, and/or modulate lactose homeostasis. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:14, or substitution, deletion or insertion of non-critical residues in non-critical regions of the polypeptide.

[0673] Non-functional allelic variants are naturally occurring amino acid sequence variants of the human HGT-1 polypeptide that do not have an HGT-1 activity, *e.g.*, they do not have the ability to bind UDP-galactose and N-acetylglucosamine, form glycosidic bonds or to modulate lactose homeostasis. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:14, or a substitution, insertion or deletion in critical residues or critical regions.

[0674] The present invention further provides non-human orthologues of the human HGT-1 polypeptide. Orthologues of human HGT-1 polypeptides are polypeptides that are isolated from non-human organisms and possess the same HGT-1 activity, *e.g.*, ligand binding, and/or modulation of galactosyltransferase activities, and/or modulation of lactose homeostasis, as the human HGT-1 polypeptide. Orthologues of the human HGT-1 polypeptide can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:14.

[0675] Moreover, nucleic acid molecules encoding other HGT-1 family members and, thus, which have a nucleotide sequence which differs from the HGT-1 sequences of SEQ ID NO:13 or 15 are intended to be within the scope of the invention. For example, another HGT-1 cDNA can be identified based on the nucleotide sequence of human HGT-1. Moreover, nucleic acid molecules encoding HGT-1 polypeptides from different species, and which, thus, have a nucleotide sequence which differs from the HGT-1 sequences of SEQ ID NO:13 or 15 are intended to be within the scope of the invention. For example, a mouse HGT-1 cDNA can be identified based on the nucleotide sequence of a human HGT-1.

[0676] Nucleic acid molecules corresponding to natural allelic variants and homologues of the HGT-1 cDNAs of the invention can be isolated based on their homology to the HGT-1 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the HGT-1 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the HGT-1 gene.

[0677] Orthologues, homologues and allelic variants can be identified using methods known in the art (*e.g.*, by hybridization to an isolated nucleic acid molecule of the present invention, for example, under stringent hybridization conditions). In one embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:13 or 15. In other embodiment, the nucleic acid is at least 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, 3050, 3100, 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, 3600, 3650, 3700, 3750, 3800, 3850, 3900, 3950, 4000 or more nucleotides in length.

[0678] As used herein, the term “hybridizes under stringent conditions” is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X

SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45°C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH₂PO₄, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, $T_m(^{\circ}\text{C}) = 2(\# \text{ of A + T bases}) + 4(\# \text{ of G + C bases})$. For hybrids between 18 and 49 base pairs in length, $T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{G+C}) - (600/\text{N})$, where N is the number of bases in the hybrid, and $[\text{Na}^+]$ is the concentration of sodium ions in the hybridization buffer ($[\text{Na}^+]$ for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH₂PO₄, 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH₂PO₄, 1% SDS at 65°C, see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (or alternatively 0.2X SSC, 1% SDS).

[0679] Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:13 or 15 and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural polypeptide).

[0680] In addition to naturally-occurring allelic variants of the HGT-1 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be

introduced by mutation into the nucleotide sequences of SEQ ID NO:13 or 15, thereby leading to changes in the amino acid sequence of the encoded HGT-1 polypeptides, without altering the functional ability of the HGT-1 polypeptides. For example, nucleotide substitutions leading to amino acid substitutions at “non-essential” amino acid residues can be made in the sequence of SEQ ID NO:13 or 15. A “non-essential” amino acid residue is a residue that can be altered from the wild-type sequence of HGT-1 (*e.g.*, the sequence of SEQ ID NO:14) without altering the biological activity, whereas an “essential” amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the HGT-1 polypeptides of the present invention, *e.g.*, those present in a transmembrane domain and/or a galactosyltransferase family domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the HGT-1 polypeptides of the present invention and other members of the HGT-1 family are not likely to be amenable to alteration.

[0681] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding HGT-1 polypeptides that contain changes in amino acid residues that are not essential for activity. Such HGT-1 polypeptides differ in amino acid sequence from SEQ ID NO:14, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to SEQ ID NO:14 (*e.g.*, to the entire length of SEQ ID NO:14).

[0682] An isolated nucleic acid molecule encoding an HGT-1 polypeptide identical to the polypeptide of SEQ ID NO:14, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:13 or 15, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded polypeptide. Mutations can be introduced into SEQ ID NO:13 or 15 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*,

aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine, tryptophan), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in an HGT-1 polypeptide is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an HGT-1 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for HGT-1 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:13 or 15, the encoded polypeptide can be expressed recombinantly and the activity of the polypeptide can be determined.

[0683] In a preferred embodiment, a mutant HGT-1 polypeptide can be assayed for the ability to (i) bind UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine bound to a glycoprotein); (ii) catalyze the formation of glycosidic bonds (*e.g.*, between UDP-galactose and N-acetylglucosamine); (iii) modulate lactose homeostasis; (iv) regulate embryogenesis; (v) regulate development; (vi) regulate the formation of structural elements of the cell; (vii) regulate the metabolism of adhesive ligands; (viii) regulate the metabolism of glycoprotein ligands and receptors; (ix) regulate blood clotting; (x) regulate thrombus dissolution; (xi) regulate hormone action; (xii) regulate fertilization; (xiii) regulate an immune system response; and/or (xiv) regulate cellular proliferation, growth, differentiation, and/or migration.

[0684] In addition to the nucleic acid molecules encoding HGT-1 polypeptides described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. In an exemplary embodiment, the invention provides an isolated nucleic acid molecule which is antisense to an HGT-1 nucleic acid molecule (*e.g.*, is antisense to the coding strand of an HGT-1 nucleic acid molecule). An “antisense” nucleic acid comprises a nucleotide sequence which is complementary to a “sense” nucleic acid encoding a polypeptide, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire HGT-1 coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a “coding region” of the coding strand of a nucleotide sequence encoding HGT-1. The term “coding region” refers to

the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human HGT-1 corresponds to SEQ ID NO:15). In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strand of a nucleotide sequence encoding HGT-1. The term “noncoding region” refers to 5’ and 3’ sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5’ and 3’ untranslated regions).

[0685] Given the coding strand sequences encoding HGT-1 disclosed herein (*e.g.*, SEQ ID NO:15), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of HGT-1 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of HGT-1 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of HGT-1 mRNA (*e.g.*, between the –10 and +10 regions of the start site of a gene nucleotide sequence). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil,

(acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0686] The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an HGT-1 polypeptide to thereby inhibit expression of the polypeptide, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[0687] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-*o*-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

[0688] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in

Haseloff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave HGT-1 mRNA transcripts to thereby inhibit translation of HGT-1 mRNA. A ribozyme having specificity for an HGT-1-encoding nucleic acid can be designed based upon the nucleotide sequence of an HGT-1 cDNA disclosed herein (*i.e.*, SEQ ID NO:13 or 15). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an HGT-1-encoding mRNA. See, *e.g.*, Cech *et al.*, U.S. Patent No. 4,987,071; and Cech *et al.*, U.S. Patent No. 5,116,742. Alternatively, HGT-1 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[0689] Alternatively, HGT-1 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the HGT-1 (*e.g.*, the HGT-1 promoter and/or enhancers) to form triple helical structures that prevent transcription of the HGT-1 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioessays* 14(12):807-15.

[0690] In yet another embodiment, the HGT-1 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup, B. and Nielsen, P.E. (1996) *Bioorg. Med. Chem.* 4(1):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup and Nielsen (1996) *supra* and Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670-675.

[0691] PNAs of HGT-1 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of HGT-1 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene (*e.g.*, by PNA-directed PCR

clamping); as 'artificial restriction enzymes' when used in combination with other enzymes (*e.g.*, S1 nucleases (Hyrup and Nielsen (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup and Nielsen (1996) *supra*; Perry-O'Keefe *et al.* (1996) *supra*).

[0692] In another embodiment, PNAs of HGT-1 can be modified (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of HGT-1 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup and Nielsen (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup and Nielsen (1996) *supra* and Finn, P. J. *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

[0693] In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may

be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0694] Alternatively, the expression characteristics of an endogenous HGT-1 gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous HGT-1 gene. For example, an endogenous HGT-1 gene which is normally “transcriptionally silent”, *i.e.*, an HGT-1 gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous HGT-1 gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

[0695] A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous HGT-1 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

Isolated HGT-1 Polypeptides and Anti-HGT-1 Antibodies

[0696] One aspect of the invention pertains to isolated HGT-1 or recombinant polypeptides and polypeptides, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-HGT-1 antibodies. In one embodiment, native HGT-1 polypeptides can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, HGT-1 polypeptides are produced by recombinant DNA techniques. Alternative to recombinant expression, an HGT-1 polypeptide or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

[0697] An “isolated” or “purified” polypeptide or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the HGT-1 polypeptide is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language “substantially free of cellular material” includes preparations of HGT-1 polypeptide in which the polypeptide is separated from cellular components of the cells from which it is

isolated or recombinantly produced. In one embodiment, the language “substantially free of cellular material” includes preparations of HGT-1 polypeptide having less than about 30% (by dry weight) of non-HGT-1 polypeptide (also referred to herein as a “contaminating protein”), more preferably less than about 20% of non-HGT-1 polypeptide, still more preferably less than about 10% of non-HGT-1 polypeptide, and most preferably less than about 5% non-HGT-1 polypeptide. When the HGT-1 polypeptide or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

[0698] The language “substantially free of chemical precursors or other chemicals” includes preparations of HGT-1 polypeptide in which the polypeptide is separated from chemical precursors or other chemicals which are involved in the synthesis of the polypeptide. In one embodiment, the language “substantially free of chemical precursors or other chemicals” includes preparations of HGT-1 polypeptide having less than about 30% (by dry weight) of chemical precursors or non-HGT-1 chemicals, more preferably less than about 20% chemical precursors or non-HGT-1 chemicals, still more preferably less than about 10% chemical precursors or non-HGT-1 chemicals, and most preferably less than about 5% chemical precursors or non-HGT-1 chemicals.

[0699] As used herein, a “biologically active portion” of an HGT-1 polypeptide includes a fragment of an HGT-1 polypeptide which participates in an interaction between an HGT-1 molecule and a non-HGT-1 molecule. Biologically active portions of an HGT-1 polypeptide include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the HGT-1 polypeptide, *e.g.*, the amino acid sequence shown in SEQ ID NO:14, which include less amino acids than the full length HGT-1 polypeptides, and exhibit at least one activity of an HGT-1 polypeptide. Typically, biologically active portions comprise a domain or motif with at least one activity of the HGT-1 polypeptide, *e.g.*, modulating galactosyltransferase activities. A biologically active portion of an HGT-1 polypeptide can be a polypeptide which is, for example, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375 or more amino acids in length. Biologically active portions of an HGT-1 polypeptide can be used as targets for developing agents which modulate an HGT-1 activity.

[0700] In one embodiment, a biologically active portion of an HGT-1 polypeptide comprises at least one transmembrane domain. It is to be understood that a preferred

biologically active portion of an HGT-1 polypeptide of the present invention comprises at least one or more of the following domains: a transmembrane domain and/or a galactosyltransferase family domain. Moreover, other biologically active portions, in which other regions of the polypeptide are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native HGT-1 polypeptide.

[0701] Another aspect of the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NO:14, for example, for use as immunogens. In one embodiment, a fragment comprises at least 5 amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:14. In another embodiment, a fragment comprises at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:14.

[0702] In a preferred embodiment, an HGT-1 polypeptide has an amino acid sequence shown in SEQ ID NO:14. In other embodiments, the HGT-1 polypeptide is substantially identical to SEQ ID NO:14, and retains the functional activity of the polypeptide of SEQ ID NO:14, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. In another embodiment, the HGT-1 polypeptide is a polypeptide which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to SEQ ID NO:14.

[0703] In another embodiment, the invention features an HGT-1 polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more identical to a nucleotide sequence of SEQ ID NO:13 or SEQ ID NO:15, or a complement thereof. This invention further features an HGT-1 polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:13 or SEQ ID NO:15, or a complement thereof.

[0704] To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for

comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the HGT-1 amino acid sequence of SEQ ID NO:14 having 378 amino acid residues, at least 113, preferably at least 151, more preferably at least 189, more preferably at least 227, even more preferably at least 265, and even more preferably at least 302 or 340 or more amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid “identity” is equivalent to amino acid or nucleic acid “homology”). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0705] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available online through the Genetics Computer Group), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available online through the Genetics Computer Group), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blosum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0706] In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of Meyers, E. and Miller, W. (*Comput. Appl. Biosci.* 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or version 2.0U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0707] The nucleic acid and polypeptide sequences of the present invention can further be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to HGT-1 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 100, wordlength = 3, and a Blosum62 matrix to obtain amino acid sequences homologous to HGT-1 polypeptide molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See the internet website for the National Center for Biotechnology Information.

[0708] The invention also provides HGT-1 chimeric or fusion proteins. As used herein, an HGT-1 “chimeric protein” or “fusion protein” comprises an HGT-1 polypeptide operatively linked to a non-HGT-1 polypeptide. An “HGT-1 polypeptide” refers to a polypeptide having an amino acid sequence corresponding to HGT-1, whereas a “non-HGT-1 polypeptide” refers to a polypeptide having an amino acid sequence corresponding to a polypeptide which is not substantially homologous to the HGT-1 polypeptide, *e.g.*, a polypeptide which is different from the HGT-1 polypeptide and which is derived from the same or a different organism. Within an HGT-1 fusion protein the HGT-1 polypeptide can correspond to all or a portion of an HGT-1 polypeptide. In a preferred embodiment, an HGT-1 fusion protein comprises at least one biologically active portion of an HGT-1 polypeptide. In another preferred embodiment, an HGT-1 fusion protein comprises at least two biologically active portions of an HGT-1 polypeptide. Within the fusion protein, the term “operatively linked” is intended to indicate that the HGT-1 polypeptide and the non-HGT-1 polypeptide are fused in-frame to each other. The non-HGT-1 polypeptide can be fused to the N-terminus or C-terminus of the HGT-1 polypeptide.

[0709] For example, in one embodiment, the fusion protein is a GST-HGT-1 fusion protein in which the HGT-1 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant HGT-1. In another embodiment, the fusion protein is an HGT-1 polypeptide containing a heterologous signal

sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of HGT-1 can be increased through the use of a heterologous signal sequence.

[0710] The HGT-1 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The HGT-1 fusion proteins can be used to affect the bioavailability of an HGT-1 substrate. Use of HGT-1 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding an HGT-1 polypeptide; (ii) mis-regulation of the HGT-1 gene; and (iii) aberrant post-translational modification of an HGT-1 polypeptide.

[0711] Moreover, the HGT-1-fusion proteins of the invention can be used as immunogens to produce anti-HGT-1 antibodies in a subject, to purify HGT-1 ligands and in screening assays to identify molecules which inhibit the interaction of HGT-1 with an HGT-1 substrate.

[0712] Preferably, an HGT-1 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.*, John Wiley & Sons:1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An HGT-1-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the HGT-1 polypeptide.

[0713] The present invention also pertains to variants of the HGT-1 polypeptides which function as either HGT-1 agonists (mimetics) or as HGT-1 antagonists. Variants of the HGT-1 polypeptides can be generated by mutagenesis, *e.g.*, discrete point mutation or

truncation of an HGT-1 polypeptide. An agonist of the HGT-1 polypeptides can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of an HGT-1 polypeptide. An antagonist of an HGT-1 polypeptide can inhibit one or more of the activities of the naturally occurring form of the HGT-1 polypeptide by, for example, competitively modulating an HGT-1-mediated activity of an HGT-1 polypeptide. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the polypeptide has fewer side effects in a subject relative to treatment with the naturally occurring form of the HGT-1 polypeptide.

[0714] In one embodiment, variants of an HGT-1 polypeptide which function as either HGT-1 agonists (mimetics) or as HGT-1 antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of an HGT-1 polypeptide for HGT-1 polypeptide agonist or antagonist activity. In one embodiment, a variegated library of HGT-1 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of HGT-1 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential HGT-1 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of HGT-1 sequences therein. There are a variety of methods which can be used to produce libraries of potential HGT-1 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential HGT-1 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acids Res.* 11:477.

[0715] In addition, libraries of fragments of an HGT-1 polypeptide coding sequence can be used to generate a variegated population of HGT-1 fragments for screening and subsequent selection of variants of an HGT-1 polypeptide. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an HGT-1 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form

double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the HGT-1 polypeptide.

[0716] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of HGT-1 polypeptides. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify HGT-1 variants (Arkin and Youvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delagrave *et al.* (1993) *Protein Eng.* 6(3):327-331).

[0717] In one embodiment, cell based assays can be exploited to analyze a variegated HGT-1 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, an endothelial cell line, which ordinarily responds to HGT-1 in a particular HGT-1 substrate-dependent manner. The transfected cells are then contacted with HGT-1 and the effect of expression of the mutant on signaling by the HGT-1 substrate can be detected, *e.g.*, by monitoring intracellular calcium, IP3, or diacylglycerol concentration, phosphorylation profile of intracellular proteins, or the activity of an HGT-1-regulated transcription factor. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the HGT-1 substrate, and the individual clones further characterized.

[0718] An isolated HGT-1 polypeptide, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind HGT-1 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length HGT-1 polypeptide can be used or, alternatively, the invention provides antigenic peptide fragments of HGT-1 for use as immunogens. The antigenic peptide of HGT-1 comprises at least 8 amino acid residues

of the amino acid sequence shown in SEQ ID NO:14 and encompasses an epitope of HGT-1 such that an antibody raised against the peptide forms a specific immune complex with HGT-1. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[0719] Preferred epitopes encompassed by the antigenic peptide are regions of HGT-1 that are located on the surface of the polypeptide, *e.g.*, hydrophilic regions, as well as regions with high antigenicity.

[0720] An HGT-1 immunogen typically is used to prepare antibodies by immunizing a suitable subject (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed HGT-1 polypeptide or a chemically synthesized HGT-1 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic HGT-1 preparation induces a polyclonal anti-HGT-1 antibody response.

[0721] Accordingly, another aspect of the invention pertains to anti-HGT-1 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as HGT-1. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind HGT-1. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of HGT-1. A monoclonal antibody composition thus typically displays a single binding affinity for a particular HGT-1 polypeptide with which it immunoreacts.

[0722] Polyclonal anti-HGT-1 antibodies can be prepared as described above by immunizing a suitable subject with an HGT-1 immunogen. The anti-HGT-1 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized HGT-1. If desired, the antibody molecules directed against HGT-1 can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to

obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-HGT-1 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem.* 255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally Kenneth, R.H. in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); Lerner, E.A. (1981) *Yale J. Biol. Med.* 54:387-402; Gefter, M.L. *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an HGT-1 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds HGT-1.

[0723] Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-HGT-1 monoclonal antibody (see, *e.g.*, Galfre, G. *et al.* (1977) *Nature* 266:55052; Gefter *et al.* (1977) *supra*; Lerner (1981) *supra*; Kenneth (1980) *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells

(unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind HGT-1, *e.g.*, using a standard ELISA assay.

[0724] Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-HGT-1 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with HGT-1 to thereby isolate immunoglobulin library members that bind HGT-1. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.*, U.S. Patent No. 5,223,409; Kang *et al.*, PCT International Publication No. WO 92/18619; Dower *et al.*, PCT International Publication No. WO 91/17271; Winter *et al.*, PCT International Publication No. WO 92/20791; Markland *et al.*, PCT International Publication No. WO 92/15679; Breitling *et al.*, PCT International Publication No. WO 93/01288; McCafferty *et al.*, PCT International Publication No. WO 92/01047; Garrard *et al.*, PCT International Publication No. WO 92/09690; Ladner *et al.*, PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Biotechnology (NY)* 9:1369-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clackson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard *et al.* (1991) *Biotechnology (NY)* 9:1373-1377; Hoogenboom *et al.* (1991) *Nucleic Acids Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al.* (1990) *Nature* 348:552-554.

[0725] Additionally, recombinant anti-HGT-1 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.*, International Application No. PCT/US86/02269; Akira *et al.*, European Patent Application No. 184,187; Taniguchi, M., European Patent Application 171,496;

Morrison *et al.*, European Patent Application 173,494; Neuberger *et al.*, PCT International Publication No. WO 86/01533; Cabilly *et al.*, U.S. Patent No. 4,816,567; Cabilly *et al.*, European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S.L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *Biotechniques* 4:214; Winter, U.S. Patent No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyen *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

[0726] An anti-HGT-1 antibody (*e.g.*, monoclonal antibody) can be used to isolate HGT-1 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-HGT-1 antibody can facilitate the purification of natural HGT-1 from cells and of recombinantly produced HGT-1 expressed in host cells. Moreover, an anti-HGT-1 antibody can be used to detect HGT-1 polypeptide (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the HGT-1 polypeptide. Anti-HGT-1 antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Recombinant Expression Vectors and Host Cells

[0727] Another aspect of the invention pertains to vectors, for example recombinant expression vectors, containing a nucleic acid containing an HGT-1 nucleic acid molecule or

vectors containing a nucleic acid molecule which encodes an HGT-1 polypeptide (or a portion thereof). As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “expression vectors”. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, “plasmid” and “vector” can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0728] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operably linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term “regulatory sequence” is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel (1990) *Methods Enzymol.* 185:3-7. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can

depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, HGT-1 polypeptides, mutant forms of HGT-1 polypeptides, fusion proteins, and the like).

[0729] Accordingly, an exemplary embodiment provides a method for producing a polypeptide, preferably an HGT-1 polypeptide, by culturing in a suitable medium a host cell of the invention (*e.g.*, a mammalian host cell such as a non-human mammalian cell) containing a recombinant expression vector, such that the polypeptide is produced.

[0730] The recombinant expression vectors of the invention can be designed for expression of HGT-1 polypeptides in prokaryotic or eukaryotic cells. For example, HGT-1 polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel (1990) *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0731] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0732] Purified fusion proteins can be utilized in HGT-1 activity assays (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for

HGT-1 polypeptides, for example. In a preferred embodiment, an HGT-1 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

[0733] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Methods Enzymol.* 185:60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid *trp-lac* fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 *gn10-lac* fusion promoter mediated by a coexpressed viral RNA polymerase (T7 *gn1*). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 *gn1* gene under the transcriptional control of the *lacUV 5* promoter.

[0734] One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S. (1990) *Methods Enzymol.* 185:119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.* (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0735] In another embodiment, the HGT-1 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.* (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.* (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).

[0736] Alternatively, HGT-1 polypeptides can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

[0737] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC

(Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

[0738] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[0739] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to HGT-1 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic

acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

[0740] Another aspect of the invention pertains to host cells into which an HGT-1 nucleic acid molecule of the invention is introduced, *e.g.*, an HGT-1 nucleic acid molecule within a vector (*e.g.*, a recombinant expression vector) or an HGT-1 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0741] A host cell can be any prokaryotic or eukaryotic cell. For example, an HGT-1 polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0742] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

[0743] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers

include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an HGT-1 polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0744] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) an HGT-1 polypeptide. Accordingly, the invention further provides methods for producing an HGT-1 polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding an HGT-1 polypeptide has been introduced) in a suitable medium such that an HGT-1 polypeptide is produced. In another embodiment, the method further comprises isolating an HGT-1 polypeptide from the medium or the host cell.

[0745] The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which HGT-1-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous HGT-1 sequences have been introduced into their genome or homologous recombinant animals in which endogenous HGT-1 sequences have been altered. Such animals are useful for studying the function and/or activity of an HGT-1 and for identifying and/or evaluating modulators of HGT-1 activity. As used herein, a “transgenic animal” is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a “homologous recombinant animal” is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous HGT-1 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced

into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

[0746] A transgenic animal of the invention can be created by introducing an HGT-1-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The HGT-1 cDNA sequence of SEQ ID NO:13 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human HGT-1 gene, such as a mouse or rat HGT-1 gene, can be used as a transgene. Alternatively, an HGT-1 gene homologue, such as another HGT-1 family member, can be isolated based on hybridization to the HGT-1 cDNA sequences of SEQ ID NO:13 or 15 (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to an HGT-1 transgene to direct expression of an HGT-1 polypeptide to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of an HGT-1 transgene in its genome and/or expression of HGT-1 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding an HGT-1 polypeptide can further be bred to other transgenic animals carrying other transgenes.

[0747] To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an HGT-1 gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the HGT-1 gene. The HGT-1 gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:15), but more preferably, is a non-human homologue of a human HGT-1 gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:13). For example, a mouse HGT-1 gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous HGT-1 gene in the mouse genome. In a

preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous HGT-1 gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a “knock out” vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous HGT-1 gene is mutated or otherwise altered but still encodes functional polypeptide (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous HGT-1 polypeptide). In the homologous recombination nucleic acid molecule, the altered portion of the HGT-1 gene is flanked at its 5’ and 3’ ends by additional nucleic acid sequence of the HGT-1 gene to allow for homologous recombination to occur between the exogenous HGT-1 gene carried by the homologous recombination nucleic acid molecule and an endogenous HGT-1 gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking HGT-1 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5’ and 3’ ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K.R. and Capecchi, M.R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced HGT-1 gene has homologously recombined with the endogenous HGT-1 gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then be injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, E.J., ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos. WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

[0748] In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One

example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O’Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of “double” transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

[0749] Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmot, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_O phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

Pharmaceutical Compositions

[0750] The HGT-1 nucleic acid molecules, fragments of HGT-1 polypeptides, and anti-HGT-1 antibodies (also referred to herein as “active compounds”) of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, polypeptide, or antibody and a pharmaceutically acceptable carrier. As used herein the language “pharmaceutically acceptable carrier” is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

[0751] A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0752] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0753] Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of an HGT-1 polypeptide or an anti-HGT-1 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0754] Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0755] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0756] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[0757] The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0758] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0759] It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

[0760] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0761] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0762] As defined herein, a therapeutically effective amount of polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a polypeptide or antibody can include a single treatment or, preferably, can include a series of treatments.

[0763] In a preferred example, a subject is treated with antibody or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

[0764] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic

compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

[0765] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0766] In certain embodiments of the invention, a modulator of HGT-1 activity is administered in combination with other agents (*e.g.*, a small molecule), or in conjunction with another, complementary treatment regime. For example, in one embodiment, a modulator of HGT-1 activity is used to treat a cellular proliferation, growth, differentiation, and/or migration disorder. Accordingly, modulation of HGT-1 activity may be used in

conjunction with, for example, another agent or treatment used to treat the disorder, *e.g.*, radiation or conventional chemotherapy.

[0767] Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologues thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

[0768] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[0769] Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting of Drugs in Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in *Controlled Drug Delivery (2nd Ed.)*, Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in

Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", *Immunol. Rev.* 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0770] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent No. 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[0771] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Uses and Methods of the Invention

[0772] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, an HGT-1 polypeptide of the invention has one or more of the following activities: (i) it may bind UDP-galactose and N-acetylglucosamine (*e.g.*, N-acetylglucosamine bound to a glycoprotein); (ii) it may catalyze the formation of glycosidic bonds (*e.g.*, between UDP-galactose and N-acetylglucosamine); (iii) it may modulate lactose homeostasis; (iv) it may regulate embryogenesis; (v) it may regulate development; (vi) it may regulate the formation of structural elements of the cell; (vii) it may regulate the metabolism of adhesive ligands; (viii) it may regulate the metabolism of glycoprotein ligands and receptors; (ix) it may regulate blood clotting; (x) it may regulate thrombus

dissolution; (xi) it may regulate hormone action; (xii) it may regulate fertilization; (xiii) it may regulate an immune system response; and/or (xiv) it may regulated cellular proliferation, growth, differentiation, and/or migration.

[0773] The isolated nucleic acid molecules of the invention can be used, for example, to express HGT-1 polypeptides (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect HGT-1 mRNA (*e.g.*, in a biological sample) or a genetic alteration in an HGT-1 gene, and to modulate HGT-1 activity, as described further below. The HGT-1 polypeptides can be used to treat disorders characterized by insufficient or excessive production of an HGT-1 substrate or production of HGT-1 inhibitors. In addition, the HGT-1 polypeptides can be used to screen for naturally occurring HGT-1 substrates, to screen for drugs or compounds which modulate HGT-1 activity, as well as to treat disorders characterized by insufficient or excessive production of HGT-1 polypeptide or production of HGT-1 polypeptide forms which have decreased, aberrant or unwanted activity compared to HGT-1 wild type polypeptide (*e.g.*, galactosyltransferase associated disorders). Moreover, the anti-HGT-1 antibodies of the invention can be used to detect and isolate HGT-1 polypeptides, to regulate the bioavailability of HGT-1 polypeptides, and modulate HGT-1 activity.

Screening Assays

[0774] The invention provides a method (also referred to herein as a “screening assay”) for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to HGT-1 polypeptides, have a stimulatory or inhibitory effect on, for example, HGT-1 expression or HGT-1 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of HGT-1 substrate.

[0775] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of an HGT-1 polypeptide or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of an HGT-1 polypeptide or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods

requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[0776] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

[0777] Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

[0778] In one embodiment, an assay is a cell-based assay in which a cell which expresses an HGT-1 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate HGT-1 activity is determined. Determining the ability of the test compound to modulate HGT-1 activity can be accomplished by monitoring, for example, intracellular or extracellular UDP-galactose, UMP-galactose, N-acetylglucosamine, or N-acetyllactosamine concentration; glycoprotein synthesis; or cellular growth or proliferation.

[0779] The ability of the test compound to modulate HGT-1 binding to a substrate or to bind to HGT-1 can also be determined. Determining the ability of the test compound to modulate HGT-1 binding to a substrate can be accomplished, for example, by coupling the HGT-1 substrate with a radioisotope or enzymatic label such that binding of the HGT-1 substrate to HGT-1 can be determined by detecting the labeled HGT-1 substrate in a complex. Alternatively, HGT-1 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate HGT-1 binding to an HGT-1 substrate in a complex. Determining the ability of the test compound to bind HGT-1 can be

accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to HGT-1 can be determined by detecting the labeled HGT-1 compound in a complex. For example, compounds (*e.g.*, HGT-1 substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[0780] It is also within the scope of this invention to determine the ability of a compound (*e.g.*, an HGT-1 substrate) to interact with HGT-1 without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with HGT-1 without the labeling of either the compound or the HGT-1.

McConnell, H.M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a “microphysiometer” (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and HGT-1.

[0781] In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing an HGT-1 target molecule (*e.g.*, an HGT-1 substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the HGT-1 target molecule. Determining the ability of the test compound to modulate the activity of an HGT-1 target molecule can be accomplished, for example, by determining the ability of the HGT-1 polypeptide to bind to or interact with the HGT-1 target molecule.

[0782] Determining the ability of the HGT-1 polypeptide, or a biologically active fragment thereof, to bind to or interact with an HGT-1 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the HGT-1 polypeptide to bind to or interact with an HGT-1 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular Ca^{2+} , diacylglycerol, IP_3 , and the like), detecting catalytic/enzymatic activity of the target using an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive

regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response.

[0783] In yet another embodiment, an assay of the present invention is a cell-free assay in which an HGT-1 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the HGT-1 polypeptide or biologically active portion thereof is determined. Preferred biologically active portions of the HGT-1 polypeptides to be used in assays of the present invention include fragments which participate in interactions with non-HGT-1 molecules, *e.g.*, fragments with high surface probability scores. Binding of the test compound to the HGT-1 polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the HGT-1 polypeptide or biologically active portion thereof with a known compound which binds HGT-1 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an HGT-1 polypeptide, wherein determining the ability of the test compound to interact with an HGT-1 polypeptide comprises determining the ability of the test compound to preferentially bind to HGT-1 or biologically active portion thereof as compared to the known compound.

[0784] In another embodiment, the assay is a cell-free assay in which an HGT-1 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the HGT-1 polypeptide or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of an HGT-1 polypeptide can be accomplished, for example, by determining the ability of the HGT-1 polypeptide to bind to an HGT-1 target molecule by one of the methods described above for determining direct binding. Determining the ability of the HGT-1 polypeptide to bind to an HGT-1 target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

[0785] In an alternative embodiment, determining the ability of the test compound to modulate the activity of an HGT-1 polypeptide can be accomplished by determining the ability of the HGT-1 polypeptide to further modulate the activity of a downstream effector of an HGT-1 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

[0786] In yet another embodiment, the cell-free assay involves contacting an HGT-1 polypeptide or biologically active portion thereof with a known compound which binds the HGT-1 polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the HGT-1 polypeptide, wherein determining the ability of the test compound to interact with the HGT-1 polypeptide comprises determining the ability of the HGT-1 polypeptide to preferentially bind to or modulate the activity of an HGT-1 target molecule.

[0787] In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either HGT-1 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to an HGT-1 polypeptide, or interaction of an HGT-1 polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ HGT-1 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized micrometer plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or HGT-1 polypeptide, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or micrometer plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of HGT-1 binding or activity determined using standard techniques.

[0788] Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either an HGT-1 polypeptide or an HGT-1 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated HGT-1 polypeptide or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with HGT-1 polypeptide or target molecules but which do not interfere with binding of the HGT-1 polypeptide to its target molecule can be derivatized to the wells of the plate, and unbound target or HGT-1 polypeptide trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the HGT-1 polypeptide or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the HGT-1 polypeptide or target molecule.

[0789] In another embodiment, modulators of HGT-1 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of HGT-1 mRNA or polypeptide in the cell is determined. The level of expression of HGT-1 mRNA or polypeptide in the presence of the candidate compound is compared to the level of expression of HGT-1 mRNA or polypeptide in the absence of the candidate compound. The candidate compound can then be identified as a modulator of HGT-1 expression based on this comparison. For example, when expression of HGT-1 mRNA or polypeptide is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of HGT-1 mRNA or polypeptide expression. Alternatively, when expression of HGT-1 mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of HGT-1 mRNA or polypeptide expression. The level of HGT-1 mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting HGT-1 mRNA or polypeptide.

[0790] In yet another aspect of the invention, the HGT-1 polypeptides can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins, which bind to

or interact with HGT-1 (“HGT-1-binding proteins” or “HGT-1-bp”) and are involved in HGT-1 activity. Such HGT-1-binding proteins are also likely to be involved in the propagation of signals by the HGT-1 polypeptides or HGT-1 targets as, for example, downstream elements of an HGT-1-mediated signaling pathway. Alternatively, such HGT-1-binding proteins are likely to be HGT-1 inhibitors.

[0791] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for an HGT-1 polypeptide is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein (“prey” or “sample”) is fused to a gene that codes for the activation domain of the known transcription factor. If the “bait” and the “prey” proteins are able to interact, *in vivo*, forming an HGT-1-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the HGT-1 polypeptide.

[0792] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of an HGT-1 polypeptide can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

[0793] For example, the ability of the agent to modulate the activity of a HGT-1 protein can be tested in an animal such as an animal model for a cellular proliferation disorder, *e.g.*, tumorigenesis. Animal based models for studying tumorigenesis *in vivo* are well known in the art (reviewed in *Animal Models of Cancer Predisposition Syndromes*, Hiai, H. and Hino, O. (eds.) 1999, *Progress in Experimental Tumor Research*, Vol. 35; Clarke, A.R. (2000) *Carcinogenesis* 21:435-41) and include, for example, carcinogen-induced tumors (Rithidech, K. *et al.* (1999) *Mutat. Res.* 428:33-39; Miller, M.L. *et al.* (2000) *Environ. Mol. Mutagen.* 35:319-327), injection and/or transplantation of tumor cells into an animal, as well as animals bearing mutations in growth regulatory genes, for example,

oncogenes (e.g., ras) (Arbeit, J.M. *et al.* (1993) *Am. J. Pathol.* 142:1187-1197; Sinn, E. *et al.* (1987) *Cell* 49:465-475; Thorgeirsson, SS *et al. Toxicol Lett* (2000) 112-113:553-555) and tumor suppressor genes (e.g., p53) (Vooijs, M. *et al.* (1999) *Oncogene* 18:5293-5303; Clark A.R. (1995) *Cancer Metast. Rev.* 14:125-148; Kumar, T.R. *et al.* (1995) *J. Intern. Med.* 238:233-238; Donehower, L.A. *et al.* (1992) *Nature* 356:215-221). Furthermore, experimental model systems are available for the study of, for example, ovarian cancer (Hamilton, T.C. *et al.* (1984) *Semin. Oncol.* 11:285-298; Rahman, N.A. *et al.* (1998) *Mol. Cell. Endocrinol.* 145:167-174; Beamer, W.G. *et al.* (1998) *Toxicol. Pathol.* 26:704-710), gastric cancer (Thompson, J. *et al.* (2000) *Int. J. Cancer* 86:863-869; Fodde, R. *et al.* (1999) *Cytogenet. Cell Genet.* 86:105-111), breast cancer (Li, M. *et al.* (2000) *Oncogene* 19:1010-1019; Green, J.E. *et al.* (2000) *Oncogene* 19:1020-1027), melanoma (Satyamoorthy, K. *et al.* (1999) *Cancer Metast. Rev.* 18:401-405), and prostate cancer (Shirai, T. *et al.* (2000) *Mutat. Res.* 462:219-226; Bostwick, D.G. *et al.* (2000) *Prostate* 43:286-294).

[0794] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., an HGT-1 modulating agent, an antisense HGT-1 nucleic acid molecule, an HGT-1-specific antibody, or an HGT-1-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

[0795] In another aspect, cell-based systems, as described herein, may be used to identify compounds which may act to ameliorate tumorigenic or proliferative disease symptoms. For example, such cell systems may be exposed to a compound, suspected of exhibiting an ability to ameliorate tumorigenic or proliferative disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of tumorigenic or proliferative disease symptoms in the exposed cells. After exposure, the cells are examined to determine whether one or more of the tumorigenic or proliferative disease cellular phenotypes has been altered to resemble a more normal or more wild type, non-tumorigenic disease or non-proliferative disease phenotype. Cellular phenotypes that are associated with tumorigenic disease states include aberrant proliferation and migration,

angiogenesis, anchorage-independent growth (i.e., attachment-independent growth), and loss of contact inhibition.

[0796] In addition, animal-based tumorigenic disease systems, such as those described herein, may be used to identify compounds capable of ameliorating tumorigenic or proliferative disease symptoms. Such animal models may be used as test substrates for the identification of drugs, pharmaceuticals, therapies, and interventions which may be effective in treating tumorigenic or proliferative disease. For example, animal models may be exposed to a compound, suspected of exhibiting an ability to ameliorate tumorigenic or proliferative disease symptoms, at a sufficient concentration and for a time sufficient to elicit such an amelioration of tumorigenic or apoptotic tumorigenic or proliferative disease symptoms in the exposed animals. The response of the animals to the exposure may be monitored by assessing the reversal of disorders associated with tumorigenic disease, for example, by counting the number of tumors and/or measuring their size before and after treatment. In addition, the animals may be monitored by assessing the reversal of disorders associated with tumorigenic disease, for example, reduction in tumor burden, tumor size, and invasive and/or metastatic potential before and after treatment.

[0797] With regard to intervention, any treatments which reverse any aspect of tumorigenic or proliferative disease symptoms should be considered as candidates for human tumorigenic or proliferative disease therapeutic intervention. Dosages of test agents may be determined by deriving dose-response curves.

[0798] Additionally, gene expression patterns may be utilized to assess the ability of a compound to ameliorate proliferative or tumorigenic disease symptoms. For example, the expression pattern of one or more genes may form part of a “gene expression profile” or “transcriptional profile” which may be then be used in such an assessment. “Gene expression profile” or “transcriptional profile”, as used herein, includes the pattern of mRNA expression obtained for a given tissue or cell type under a given set of conditions. Such conditions may include, but are not limited to, the presence of a tumor, *e.g.*, a breast or lung tumor or any of the other tumors described herein, including any of control or experimental conditions described herein.

[0799] Other conditions may include, for example, cell differentiation, transformation, metastasis, and carcinogen exposure. Gene expression profiles may be generated, for example, by utilizing a differential display procedure, Northern analysis

and/or RT-PCR. In one embodiment, HGT-1 gene sequences may be used as probes and/or PCR primers for the generation and corroboration of such gene expression profiles.

[0800] Gene expression profiles may be characterized for known states, either tumorigenic or proliferative disease or normal, within the cell- and/or animal-based model systems. Subsequently, these known gene expression profiles may be compared to ascertain the effect a test compound has to modify such gene expression profiles, and to cause the profile to more closely resemble that of a more desirable profile.

[0801] For example, administration of a compound may cause the gene expression profile of a tumorigenic or proliferative disease model system to more closely resemble the control system. Administration of a compound may, alternatively, cause the gene expression profile of a control system to begin to mimic a tumorigenic or proliferative disease state. Such a compound may, for example, be used in further characterizing the compound of interest, or may be used in the generation of additional animal models.

Detection Assays

[0802] Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[0803] Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the HGT-1 nucleotide sequences, described herein, can be used to map the location of the HGT-1 genes on a chromosome. The mapping of the HGT-1 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

[0804] Briefly, HGT-1 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the HGT-1 nucleotide sequences. Computer analysis of the HGT-1 sequences can be used to predict primers that do not span more than

one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the HGT-1 sequences will yield an amplified fragment.

[0805] Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

[0806] PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the HGT-1 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map an HGT-1 sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

[0807] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple

detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

[0808] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[0809] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data (such data are found, for example, in McKusick, V., Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature* 325:783-787.

[0810] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the HGT-1 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[0811] The HGT-1 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost,

switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent No. 5,272,057).

[0812] Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the HGT-1 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

[0813] Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The HGT-1 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:13 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:15 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[0814] If a panel of reagents from HGT-1 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of HGT-1 Sequences in Forensic Biology

[0815] DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found

at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[0816] The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another “identification marker” (*i.e.*, another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:13 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the HGT-1 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:13 having a length of at least 20 bases, preferably at least 30 bases.

[0817] The HGT-1 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such HGT-1 probes can be used to identify tissue by species and/or by organ type.

[0818] In a similar fashion, these reagents, *e.g.*, HGT-1 primers or probes can be used to screen tissue culture for contamination (*i.e.*, screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine:

[0819] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining

HGT-1 polypeptide and/or nucleic acid expression as well as HGT-1 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted HGT-1 expression or activity (*e.g.*, a cellular proliferation, growth, differentiation, or migration disorder). The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with HGT-1 polypeptide, nucleic acid expression or activity (*e.g.*, a cellular proliferation, growth, differentiation, or migration disorder). For example, mutations in an HGT-1 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with HGT-1 polypeptide, nucleic acid expression or activity.

[0820] Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of HGT-1 in clinical trials.

[0821] These and other agents are described in further detail in the following sections.

Diagnostic Assays

[0822] An exemplary method for detecting the presence or absence of HGT-1 polypeptide or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting HGT-1 polypeptide or nucleic acid (*e.g.*, mRNA, or genomic DNA) that encodes HGT-1 polypeptide such that the presence of HGT-1 polypeptide or nucleic acid is detected in the biological sample. In another aspect, the present invention provides a method for detecting the presence of HGT-1 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of HGT-1 activity such that the presence of HGT-1 activity is detected in the biological sample. A preferred agent for detecting HGT-1 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to HGT-1 mRNA or genomic DNA. The nucleic acid probe can be, for example, the HGT-1 nucleic acid set forth in SEQ ID NO:13 or 15, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to HGT-1 mRNA or genomic

DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

[0823] A preferred agent for detecting HGT-1 polypeptide is an antibody capable of binding to HGT-1 polypeptide, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term “labeled”, with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term “biological sample” is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect HGT-1 mRNA, polypeptide, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of HGT-1 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of HGT-1 polypeptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of HGT-1 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of HGT-1 polypeptide include introducing into a subject a labeled anti-HGT-1 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[0824] The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding an HGT-1 polypeptide; (ii) aberrant expression of a gene encoding an HGT-1 polypeptide; (iii) mis-regulation of the gene; and (iii) aberrant post-translational modification of an HGT-1 polypeptide, wherein a wild-type form of the gene encodes a polypeptide with an HGT-1 activity. “Misexpression or aberrant expression”, as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes, but is not limited to, expression at non-wild type levels (*e.g.*, over or under expression); a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed (*e.g.*, increased or decreased expression (as

compared with wild type) at a predetermined developmental period or stage); a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene (*e.g.*, a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus).

[0825] In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

[0826] In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting HGT-1 polypeptide, mRNA, or genomic DNA, such that the presence of HGT-1 polypeptide, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of HGT-1 polypeptide, mRNA or genomic DNA in the control sample with the presence of HGT-1 polypeptide, mRNA or genomic DNA in the test sample.

[0827] The invention also encompasses kits for detecting the presence of HGT-1 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting HGT-1 polypeptide or mRNA in a biological sample; means for determining the amount of HGT-1 in the sample; and means for comparing the amount of HGT-1 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect HGT-1 polypeptide or nucleic acid.

Prognostic Assays

[0828] The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted HGT-1 expression or activity (*e.g.*, a cellular proliferation, growth, differentiation, or migration disorder). As used herein, the term “aberrant” includes an

HGT-1 expression or activity which deviates from the wild type HGT-1 expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant HGT-1 expression or activity is intended to include the cases in which a mutation in the HGT-1 gene causes the HGT-1 gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional HGT-1 polypeptide or a polypeptide which does not function in a wild-type fashion, *e.g.*, a polypeptide which does not interact with an HGT-1 substrate, *e.g.*, a galactosyltransferase subunit or ligand, or one which interacts with a non-HGT-1 substrate, *e.g.*, a non- galactosyltransferase subunit or ligand. As used herein, the term “unwanted” includes an unwanted phenomenon involved in a biological response, such as cellular proliferation. For example, the term unwanted includes an HGT-1 expression or activity which is undesirable in a subject.

[0829] The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in HGT-1 polypeptide activity or nucleic acid expression, such as a galactosyltransferase disorder, *e.g.*, a cellular proliferation, growth, differentiation, or migration disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in HGT-1 polypeptide activity or nucleic acid expression, such as a galactosyltransferase disorder, *e.g.*, a cellular proliferation, growth, differentiation, or migration disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted HGT-1 expression or activity in which a test sample is obtained from a subject and HGT-1 polypeptide or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of HGT-1 polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted HGT-1 expression or activity. As used herein, a “test sample” refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

[0830] Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted HGT-1 expression or activity, *e.g.*, a cellular

proliferation, growth, differentiation, or migration disorder. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a galactosyltransferase disorder, *e.g.*, a cellular proliferation, growth, differentiation, or migration disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted HGT-1 expression or activity in which a test sample is obtained and HGT-1 polypeptide or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of HGT-1 polypeptide or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted HGT-1 expression or activity).

[0831] The methods of the invention can also be used to detect genetic alterations in an HGT-1 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in HGT-1 polypeptide activity or nucleic acid expression, such as a galactosyltransferase disorder, a lactose homeostasis disorder, or a disorder of cellular growth, differentiation, or migration. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding an HGT-1 -polypeptide, or the mis-expression of the HGT-1 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from an HGT-1 gene; 2) an addition of one or more nucleotides to an HGT-1 gene; 3) a substitution of one or more nucleotides of an HGT-1 gene; 4) a chromosomal rearrangement of an HGT-1 gene; 5) an alteration in the level of a messenger RNA transcript of an HGT-1 gene; 6) aberrant modification of an HGT-1 gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an HGT-1 gene; 8) a non-wild type level of an HGT-1-polypeptide; 9) allelic loss of an HGT-1 gene, and 10) inappropriate post-translational modification of an HGT-1-polypeptide. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in an HGT-1 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

[0832] In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain

reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the HGT-1-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to an HGT-1 gene under conditions such that hybridization and amplification of the HGT-1-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

[0833] Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

[0834] In an alternative embodiment, mutations in an HGT-1 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[0835] In other embodiments, genetic mutations in HGT-1 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Hum. Mutat.* 7:244-255; Kozal, M.J. *et al.* (1996) *Nat. Med.* 2:753-759). For example, genetic mutations in HGT-1 can be identified in two dimensional arrays containing light-

generated DNA probes as described in Cronin *et al.* (1996) *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[0836] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the HGT-1 gene and detect mutations by comparing the sequence of the sample HGT-1 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

[0837] Other methods for detecting mutations in the HGT-1 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type HGT-1 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.*

217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

[0838] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called “DNA mismatch repair” enzymes) in defined systems for detecting and mapping point mutations in HGT-1 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on an HGT-1 sequence, *e.g.*, a wild-type HGT-1 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

[0839] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in HGT-1 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control HGT-1 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

[0840] In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a

denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[0841] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

[0842] Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[0843] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an HGT-1 gene.

[0844] Furthermore, any cell type or tissue in which HGT-1 is expressed may be utilized in the prognostic assays described herein.

Monitoring of Effects During Clinical Trials

[0845] Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of an HGT-1 polypeptide (*e.g.*, the modulation of galactosyltransferase activity) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase HGT-1 gene expression, polypeptide levels, or upregulate HGT-1 activity, can be monitored in clinical trials of subjects exhibiting decreased HGT-1 gene expression, polypeptide levels, or downregulated HGT-1 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease HGT-1 gene expression, polypeptide levels, or downregulate HGT-1 activity, can be monitored in clinical trials of subjects exhibiting increased HGT-1 gene expression, polypeptide levels, or upregulated HGT-1 activity. In such clinical trials, the expression or activity of an HGT-1 gene, and preferably, other genes that have been implicated in, for example, an HGT-1-associated disorder can be used as a “read out” or markers of the phenotype of a particular cell.

[0846] For example, and not by way of limitation, genes, including HGT-1, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates HGT-1 activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on HGT-1-associated disorders (*e.g.*, disorders characterized by deregulated signaling or galactosyltransferase activity, *e.g.*, cellular proliferation, growth, differentiation, or migration disorders), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of HGT-1 and other genes implicated in the HGT-1-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of polypeptide produced, by one of the methods as described herein, or by measuring the levels of activity of HGT-1 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

[0847] In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i)

obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an HGT-1 polypeptide, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the HGT-1 polypeptide, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the HGT-1 polypeptide, mRNA, or genomic DNA in the pre-administration sample with the HGT-1 polypeptide, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of HGT-1 to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of HGT-1 to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent. According to such an embodiment, HGT-1 expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

Methods of Treatment:

[0848] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted HGT-1 expression or activity, *e.g.*, a galactosyltransferase associated disorder (*e.g.*, a cellular proliferation, growth, differentiation, or migration disorder). As used herein, “treatment” of a subject includes the application or administration of a therapeutic agent to a subject, or application or administration of a therapeutic agent to a cell or tissue from a subject, who has a disease or disorder, has a symptom of a disease or disorder, or is at risk of (or susceptible to) a disease or disorder, with the purpose of curing, healing, alleviating, relieving, altering, remedying, ameliorating, improving, or affecting the disease or disorder, the symptom of the disease or disorder, or the risk of (or susceptibility to) the disease or disorder. As used herein, a “therapeutic agent” includes, but is not limited to, small molecules, peptides, polypeptides, antibodies, ribozymes, and antisense oligonucleotides.

[0849] With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. “Pharmacogenomics”, as used herein, refers to the application

of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the HGT-1 molecules of the present invention or HGT-1 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

Prophylactic Methods

[0850] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted HGT-1 expression or activity, by administering to the subject an HGT-1 or an agent which modulates HGT-1 expression or at least one HGT-1 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted HGT-1 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the HGT-1 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of HGT-1 aberrancy, for example, an HGT-1, HGT-1 agonist or HGT-1 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Therapeutic Methods

[0851] Another aspect of the invention pertains to methods of modulating HGT-1 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell capable of expressing HGT-1 with an agent that modulates one or more of the activities of HGT-1 polypeptide activity associated with the cell, such that HGT-1 activity in the cell is modulated. An agent that modulates HGT-1 polypeptide activity can be an agent as described herein, such as a nucleic acid or a polypeptide, a naturally-occurring target molecule of an HGT-1 polypeptide (*e.g.*, an HGT-1 substrate), an HGT-1 antibody, an HGT-1 agonist or

antagonist, a peptidomimetic of an HGT-1 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more HGT-1 activities. Examples of such stimulatory agents include active HGT-1 polypeptide and a nucleic acid molecule encoding HGT-1 that has been introduced into the cell. In another embodiment, the agent inhibits one or more HGT-1 activities. Examples of such inhibitory agents include antisense HGT-1 nucleic acid molecules, anti-HGT-1 antibodies, and HGT-1 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of an HGT-1 polypeptide or nucleic acid molecule, e.g., a cellular proliferation, growth, differentiation, or migration disorder. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) HGT-1 expression or activity. In another embodiment, the method involves administering an HGT-1 polypeptide or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted HGT-1 expression or activity.

[0852] Stimulation of HGT-1 activity is desirable in situations in which HGT-1 is abnormally downregulated and/or in which increased HGT-1 activity is likely to have a beneficial effect. Likewise, inhibition of HGT-1 activity is desirable in situations in which HGT-1 is abnormally upregulated and/or in which decreased HGT-1 activity is likely to have a beneficial effect.

Pharmacogenomics

[0853] The HGT-1 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on HGT-1 activity (e.g., HGT-1 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) HGT-1-associated disorders (e.g., cellular proliferation, growth, differentiation, or migration disorders) associated with aberrant or unwanted HGT-1 activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug.

Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an HGT-1 molecule or HGT-1 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an HGT-1 molecule or HGT-1 modulator.

[0854] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[0855] One pharmacogenomics approach to identifying genes that predict drug response, known as “a genome-wide association”, relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a “bi-allelic” gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a “SNP” is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[0856] Alternatively, a method termed the “candidate gene approach”, can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug target is known (*e.g.*, an HGT-1 polypeptide of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[0857] As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

[0858] Alternatively, a method termed the “gene expression profiling”, can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, an HGT-1 molecule or HGT-1 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[0859] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an HGT-1 molecule or

HGT-1 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Use of HGT-1 Molecules as Surrogate Markers

[0860] The HGT-1 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the HGT-1 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the HGT-1 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a “surrogate marker” is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35:258-264; and James (1994) *AIDS Treatment News Archive* 209.

[0861] The HGT-1 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a “pharmacodynamic marker” is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to

the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (*e.g.*, an HGT-1 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-HGT-1 antibodies may be employed in an immune-based detection system for an HGT-1 polypeptide marker, or HGT-1-specific radiolabeled probes may be used to detect an HGT-1 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.*, U.S. Patent No. 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90:229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3:S16-S20.

[0862] The HGT-1 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, *e.g.*, McLeod *et al.* (1999) *Eur. J. Cancer* 35(12):1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or polypeptide (*e.g.*, HGT-1 polypeptide or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in HGT-1 DNA may correlate HGT-1 drug response. The use of

pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Electronic Apparatus Readable Media and Arrays

[0863] Electronic apparatus readable media comprising HGT-1 sequence information is also provided. As used herein, “HGT-1 sequence information” refers to any nucleotide and/or amino acid sequence information particular to the HGT-1 molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information “related to” said HGT-1 sequence information includes detection of the presence or absence of a sequence (*e.g.*, detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (*e.g.*, detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (*e.g.*, detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, “electronic apparatus readable media” refers to any suitable medium for storing, holding, or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact discs; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon HGT-1 sequence information of the present invention.

[0864] As used herein, the term “electronic apparatus” is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatuses; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

[0865] As used herein, “recorded” refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can

readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the HGT-1 sequence information.

[0866] A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, represented in the form of an ASCII file, or stored in a database application, such as DB2, Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the HGT-1 sequence information.

[0867] By providing HGT-1 sequence information in readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[0868] The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a HGT-1 associated disease or disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder, wherein the method comprises the steps of determining HGT-1 sequence information associated with the subject and based on the HGT-1 sequence information, determining whether the subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

[0869] The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder wherein the method comprises the steps of determining HGT-1 sequence information associated with the subject, and based on the HGT-1 sequence information, determining whether the subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder, and/or recommending a

particular treatment for the disease, disorder or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

[0870] The present invention also provides in a network, a method for determining whether a subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder associated with HGT-1, said method comprising the steps of receiving HGT-1 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to HGT-1 and/or a cellular proliferation, growth, differentiation, and/or migration disorder, and based on one or more of the phenotypic information, the HGT-1 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

[0871] The present invention also provides a business method for determining whether a subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder, said method comprising the steps of receiving information related to HGT-1 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to HGT-1 and/or related to a cellular proliferation, growth, differentiation, and/or migration disorder, and based on one or more of the phenotypic information, the HGT-1 information, and the acquired information, determining whether the subject has a cellular proliferation, growth, differentiation, and/or migration disorder or a pre-disposition to a cellular proliferation, growth, differentiation, and/or migration disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

[0872] The invention also includes an array comprising a HGT-1 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to

ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be HGT-1. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

[0873] In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[0874] In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a cellular proliferation, growth, differentiation, and/or migration disorder, progression of a cellular proliferation, growth, differentiation, and/or migration disorder, and processes, such a cellular transformation associated with the cellular proliferation, growth, differentiation, and/or migration disorder.

[0875] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of HGT-1 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[0876] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*,

including HGT-1) that could serve as a molecular target for diagnosis or therapeutic intervention.

[0877] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Sequence Listing, are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN HGT-1 cDNA

[0878] In this example, the identification and characterization of the gene encoding human HGT-1 (clone 8797) is described.

Isolation of the human HGT-1 cDNA

[0879] The invention is based, at least in part, on the discovery of a human gene encoding a novel polypeptide, referred to herein as human HGT-1. The entire sequence of the human clone 8797 was determined and found to contain an open reading frame termed human "HGT-1." The nucleotide sequence of the human HGT-1 gene is set forth in the Sequence Listing as SEQ ID NO:13. The amino acid sequence of the human HGT-1 expression product is set forth in the Sequence Listing as SEQ ID NO:14. The HGT-1 polypeptide comprises 378 amino acids. The coding region (open reading frame) of SEQ ID NO:13 is set forth as SEQ ID NO:15.

Analysis of the Human HGT-1 Molecules

[0880] A search using the polypeptide sequence of SEQ ID NO:14 was performed against the HMM database in PFAM resulting in the identification of a galactosyltransferase family domain in the amino acid sequence of human HGT-1 at about residues 102-321 of SEQ ID NO:14.

[0881] The amino acid sequence of human HGT-1 was analyzed using the program PSORT (available online; see Nakai, K. and Kanehisa, M. (1992) *Genomics* 14:897-911) to

predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of this analysis show that human HGT-1 may be localized to the mitochondria, cytoplasm, or Golgi complex, and has a low probability of localization in the vacuole, secretory vesicles, nucleus, and endoplasmic reticulum.

[0882] Searches of the amino acid sequence of human HGT-1 were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human HGT-1 of a number of potential N-glycosylation sites, a potential glycosaminoglycan attachment site, a number of potential protein kinase C phosphorylation sites, a number of potential casein kinase II phosphorylation sites, a potential tyrosine kinase phosphorylation site, a number of potential N-myristoylation sites, and a potential amidation site.

[0883] A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:14 was also performed, predicting one transmembrane domain in the amino acid sequence of human HGT-1 (SEQ ID NO:14) at about residues 15-32.

EXAMPLE 2: EXPRESSION OF RECOMBINANT HGT-1 POLYPEPTIDE IN BACTERIAL CELLS

[0884] In this example, human HGT-1 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, HGT-1 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-HGT-1 fusion polypeptide in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 3: EXPRESSION OF RECOMBINANT HGT-1 POLYPEPTIDE IN COS CELLS

[0885] To express the human HGT-1 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA

fragment encoding the entire HGT-1 polypeptide and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant polypeptide under the control of the CMV promoter.

[0886] To construct the plasmid, the human HGT-1 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the HGT-1 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the HGT-1 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the HGT-1 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

[0887] COS cells are subsequently transfected with the human HGT-1-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the IC54420 polypeptide is detected by radiolabeling (^{35}S -methionine or ^{35}S -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S -methionine (or ^{35}S -cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

[0888] Alternatively, DNA containing the human HGT-1 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the HGT-1 polypeptide is detected by radiolabeling and immunoprecipitation using an HGT-1-specific monoclonal antibody.

EXAMPLE 4: ANALYSIS OF HUMAN HGT-1 EXPRESSION

[0889] This example describes the expression of human HGT-1 mRNA in various tissues, tumors, cell lines, and disease models, as determined using the TaqMan™ procedure and *in situ* hybridization analysis.

In situ analysis

[0890] For *in situ* analysis, various tissues, *e.g.*, tissues obtained from lung or breast, are first frozen on dry ice. Ten-micrometer-thick sections of the tissues are postfixed with 4% formaldehyde in DEPC treated 1X phosphate-buffered saline (PBS) at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections are rinsed in DEPC 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissues are then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

[0891] Hybridizations are performed with ³⁵S-radiolabeled (5×10^7 cpm/ml) cRNA probes. Probes are incubated in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

[0892] After hybridization, slides are washed with 2X SSC. Sections were then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10 µg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides are then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections are then dehydrated rapidly through serial

ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

Taqman analysis

[0893] The Taqman™ procedure is a quantitative, real-time PCR-based approach to detecting mRNA. The RT-PCR reaction exploits the 5' nuclease activity of AmpliTaq Gold™ DNA Polymerase to cleave a TaqMan™ probe during PCR. Briefly, cDNA was generated from the samples of interest and served as the starting material for PCR amplification. In addition to the 5' and 3' gene-specific primers, a gene-specific oligonucleotide probe (complementary to the region being amplified) was included in the reaction (*i.e.*, the Taqman™ probe). The TaqMan™ probe included an oligonucleotide with a fluorescent reporter dye covalently linked to the 5' end of the probe (such as FAM (6-carboxyfluorescein), TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein), JOE (6-carboxy-4,5-dichloro-2,7-dimethoxyfluorescein), or VIC) and a quencher dye (TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine) at the 3' end of the probe.

[0894] During the PCR reaction, cleavage of the probe separated the reporter dye and the quencher dye, resulting in increased fluorescence of the reporter. Accumulation of PCR products was detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe was intact, the proximity of the reporter dye to the quencher dye resulted in suppression of the reporter fluorescence. During PCR, if the target of interest was present, the probe specifically annealed between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the AmpliTaq™ Gold DNA Polymerase cleaved the probe between the reporter and the quencher only if the probe hybridized to the target. The probe fragments were then displaced from the target, and polymerization of the strand continued. The 3' end of the probe was blocked to prevent extension of the probe during PCR. This process occurred in every cycle and did not interfere with the exponential accumulation of product. RNA was prepared using the trizol method and treated with DNase to remove contaminating genomic DNA. cDNA was synthesized using standard techniques. Mock cDNA synthesis in the absence of reverse transcriptase resulted in samples with no detectable PCR amplification of the control GAPDH or β -actin gene confirming efficient removal of genomic DNA contamination.

[0895] The expression of human HGT-1 was examined, using Taqman analysis, in various human tumors and normal human tissues. Human HGT-1 was highly expressed in coronary smooth muscle cells, static human umbilical vein endothelial cells (HUVECs), HUVECs under conditions of shear stress, kidney, skeletal muscle, normal brain cortex, prostate epithelial cells, colon tumor, and lung tumor. Expression of HGT-1 was increased in HUVECs under conditions of shear stress, as compared to static HUVECs; decreased in the heart in congestive heart failure, as compared to normal heart; increased in breast tumor, as compared to normal breast; increased in colon tumor, as compared to normal colon; and increased in lung tumor, as compared to normal lung.

[0896] The expression of human HGT-1 was further examined, using Taqman analysis, in various human tumors. Expression of human HGT-1 is increased in 4/6 breast tumors, as compared to normal breast. Human HGT-1 is also increased in 7/7 lung tumors, as compared to normal lung. Human HGT-1 is also increased in 1/4 colon tumors, as compared to normal colon, and in 1/2 colon tumor metastases to the liver, as compared to normal liver or normal colon.

[0897] The expression of human HGT-1 was further examined, using Taqman analysis, in various lung cancer models. High expression was observed in H522 adenocarcinoma (AC) cells, H520 squamous cell carcinoma (SCC) cells, H69 small cell lung cancer (SCLC) cells, and H345 (undifferentiated small cell lung cancer) cells.

[0898] Finally, the expression of human HGT-1 was examined, using Taqman analysis, in various breast cancer models. Expression of human HGT-1 is induced upon treatment of MCF10A cells with the growth factors EGF or IGF1A. MCF10A cells are immortalized, but otherwise normal cells which grow as attached cells. Expression of human HGT-1 is strongly induced in MCF10AT cells grown in Agar, as compared to MCF10AT cells grown on plastic. MCF10AT cells are pre-malignant cells with the potential for neoplastic progression (MCF10AT cells generate carcinomas in approximately 25% of xenografts). Because only neoplastic cells are capable of losing attachment-dependant growth and growing in agar, increased expression of HGT-1 in MCF10AT cells growing in agar indicates that HGT-1 expression is increased upon progression from a pre-malignant to a malignant state. Human HGT-1 expression is also increased in MCF10CA (malignant) cells grown in agar, as compared to MCF10CA cells grown on plastic.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

IV. 27439, A NOVEL HUMAN HYDROXYLASE AND USES THEREFOR

Background of the Invention

[0899] The cytochrome P450 family is a group of monooxygenase enzymes that function in the metabolism of exogenous and endogenous compounds. The mechanism by which cytochrome P450 acts involves oxidation of the target compound in order to increase water solubility, thus making the metabolized compound more readily excreted in the urine or bile. These enzymes are found primarily in microsomes in the liver, but can also be found in the intestines, lungs and other organs. At least 14 gene families have been discovered with 26 subfamilies, 20 of these mapped to the human genome. The cytochrome P450 systems fall into two major classes: bacterial/mitochondrial, which have three main components, and microsomal, which have two main components.

[0900] The main physiological action of cytochrome P450 is the addition of an oxygen atom to an organic compound. The enzyme utilizes molecular oxygen as well as hydrogen donated from NADPH in order to attach a hydroxyl group (-OH) to an alkanyl group on the target molecule. Because the function of the monooxygenases is the incorporation of a hydroxyl group, these enzymes are sometimes referred to as hydroxylases. The specificity of the reaction is dependent upon the binding site within each particular cytochrome P450 subfamily. The iron of the heme group within the enzyme is the site of the reaction and is responsible for the strong absorbance at 450 nm upon the addition of carbon monoxide, for which the family of enzymes was named.

[0901] Each isoform of the enzyme contains a structurally unique binding site in order to target different compounds. This does not preclude a particular isoform from binding multiple targets and this is often the case. Poor performing enzymes can be the result of poor specificity and binding with an isoform. For example, a patient who suffered cardiotoxicity induced by desipramine was found to be a poor metabolizer of the drug.

Alternatively, the target compound may have a competing foreign compound with a higher binding affinity. This situation may result in drug toxicity if there is a high concentration of unmetabolized drug. For example, one particular Cytochrome P450 family member, Cyp1A2, metabolizes caffeine and theophylline. The addition of the antibiotic erythromycin results in a competitive inhibition of caffeine and theophylline, and if the drugs are taken in combination can result in adverse effects such as tachycardia.

Summary of the Invention

[0902] The present invention is based, in part, on the discovery of a novel human hydroxylase, referred to herein as “27439”. The nucleotide sequence of a cDNA encoding 27439 is shown in SEQ ID NO:16, and the amino acid sequence of a 27439 polypeptide is shown in SEQ ID NO:17. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:18.

[0903] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 27439 protein or polypeptide, e.g., a biologically active portion of the 27439 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:17. In other embodiments, the invention provides an isolated 27439 nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:16, SEQ ID NO:18, wherein the nucleic acid encodes a full length 27439 protein or an active fragment thereof.

[0904] In a related aspect, the invention further provides nucleic acid constructs which include a 27439 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included, are vectors and host cells containing the 27439 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing 27439 nucleic acid molecules and polypeptides.

[0905] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 27439-encoding nucleic acids.

[0906] In still another related aspect, isolated nucleic acid molecules that are antisense to a 27439 encoding nucleic acid molecule are provided.

[0907] In another aspect, the invention features, 27439 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 27439-mediated or -related disorders. In another embodiment, the invention provides 27439 polypeptides having a 27439 activity. Preferred polypeptides are 27439 proteins including at least one Cytochrome P450 family domain, and, preferably, having a 27439 activity, e.g., a 27439 activity as described herein.

[0908] In other embodiments, the invention provides 27439 polypeptides, e.g., a 27439 polypeptide having the amino acid sequence shown in SEQ ID NO:17; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:17; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:16, SEQ ID NO:18, wherein the nucleic acid encodes a full length 27439 protein or an active fragment thereof.

[0909] In a related aspect, the invention further provides nucleic acid constructs which include a 27439 nucleic acid molecule described herein.

[0910] In a related aspect, the invention provides 27439 polypeptides or fragments operatively linked to non-27439 polypeptides to form fusion proteins.

[0911] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically bind 27439 polypeptides.

[0912] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 27439 polypeptides or nucleic acids.

[0913] In still another aspect, the invention provides a process for modulating 27439 polypeptide or nucleic acid expression or activity, e.g. using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 27439 polypeptides or nucleic acids, such as conditions involving aberrant or deficient cellular proliferation or differentiation.

[0914] The invention also provides assays for determining the activity of or the presence or absence of 27439 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[0915] In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 27439 polypeptide or nucleic acid molecule, including for disease diagnosis.

[0916] In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, *e.g.*, a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 27439 molecule. In one embodiment, the capture probe is a nucleic acid, *e.g.*, a probe complementary to a 27439 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, *e.g.*, an antibody specific for 27439 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

[0917] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Detailed Description

Human 27439

[0918] The human 27439 sequence (SEQ ID NO:16), which is approximately 1976 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1614 nucleotides (nucleotides 1-1614 of SEQ ID NO:16; SEQ ID NO:18), including the terminal codon. The coding sequence encodes a 537 amino acid protein (SEQ ID NO:17). The human 27439 protein of SEQ ID NO:17 may include a predicted amino-terminal hydrophobic amino acid sequence, consistent with a signal sequence, of about 14 amino acids (from amino acid 1 to about amino acid 14 of SEQ ID NO:17, PSORT, Nakai, K. and Kanehisa, M. (1992) *Genomics* 14:897-911), which upon cleavage results in the production of a mature protein form.

[0919] A hydropathy plot of human 27439 can be performed. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, *e.g.*, the sequence from about amino acid 1 to 8, from about 318 to 328, and from about 490 to 498 of SEQ ID NO:17; all or part of a hydrophilic sequence, *e.g.*, the sequence from about amino

acid 45 to 60, from about 135 to 150, and from about 440 to 460 of SEQ ID NO:17; a sequence which includes a Cys, or a glycosylation site.

[0920] The protein form is approximately 537 amino acid residues in length (from about amino acid 1 to amino acid 537 of SEQ ID NO:17). Alternatively, the mature protein form is approximately 523 amino acid residues in length (from about amino acid 1 to amino acid 523 of SEQ ID NO:17). Human 27439 contains the following regions or other structural features: a Cytochrome P450 family domain located at about amino acid residues 69-533; two predicted transmembrane domains which extend from about amino acid residues 316-332 and 342-362 of SEQ ID NO:17 (alternatively, the mature peptide may contain two predicted transmembrane domains which extend from about amino acid residues 302-318 and 328-348 of SEQ ID NO:17); one predicted N-glycosylation site (PS00001) located at about amino acid residues 338-341 of SEQ ID NO:17; one predicted cAMP- and cGMP-dependent protein kinase phosphorylation site (PS00004) located at about amino acid residues 314-317 of SEQ ID NO:17; two predicted protein kinase C phosphorylation sites (PS00005) located at about amino acid residues 141-143 and 403-405 of SEQ ID NO:17; seven predicted located at about amino acid residues 116-119, 141-144, 203-206, 331-334, 386-389, 417-420 and 442-445 of SEQ ID NO:17; two predicted tyrosine kinase phosphorylation sites (PS00007) located at about amino acid residues 176-183 and 302-308 of SEQ ID NO:17; eight predicted N-myristoylation sites (PS00008) located at about amino acid residues 19-24, 52-57, 154-159, 236-241, 318-323, 346-351, 431-436 and 479-484 of SEQ ID NO:17; two predicted amidation sites (PS00009) located at about amino acid residues 312-315 and 484-487 of SEQ ID NO:17; and one cytochrome P450 cysteine heme-iron ligand signature site (PS00086) located at about amino acid residues 476-485 of SEQ ID NO:17.

[0921] For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997) *Protein* 28:405-420.

[0922] The 27439 protein contains a significant number of structural characteristics in common with members of the Cytochrome P450 family. The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well

as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

[0923] As used herein, the term "Cytochrome P450 family" includes a molecule which is involved in the enzymatic hydroxylation associated with a large number of hydrophobic compounds. Cytochrome P450 family molecules are involved in the growth, development, and differentiation of cells, in the regulation of cellular homeostasis, in the metabolism and catabolism of biochemical molecules necessary for energy production or storage, in intra- or intercellular signaling, in metabolism or catabolism of metabolically important biomolecules, and in the removal of potentially harmful compounds from the interior of the cell. Examples of Cytochrome P450 family members include Cyp2D6, Cyp3A and Cyp1A2. The Cytochrome P450 family molecules of the present invention provide novel diagnostic targets and therapeutic agents to control Cytochrome P450 family-associated disorders.

[0924] As used herein, a "27439 activity", "biological activity of 27439" or "functional activity of 27439", refers to an activity exerted by a 27439 protein, polypeptide or nucleic acid molecule on e.g., a 27439-responsive cell or on a 27439 substrate, e.g., a lipid or protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 27439 activity is a direct activity, such as an association with a 27439 target molecule. A "target molecule" or "binding partner" is a molecule with which a 27439 protein binds or interacts in nature, e.g., a molecule in which the 27439 protein enzymatically attaches a hydroxyl group. A 27439 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 27439 protein with a 27439 ligand. For example, the 27439 proteins of the present invention can have one or more of the following activities: 1) association of a substrate or binding partner; 2) chemical modification of substrate or binding partner via addition of side chain group; 3) modulation of substrate or binding partner hydrophilicity; 4) modulation of foreign drug concentration; and 5) the ability to antagonize or inhibit, competitively or non-competitively, any of 1-4. Thus, the 27439 molecules can act as novel diagnostic targets and therapeutic agents for controlling Cytochrome P450-related disorders, for example, such as those diseases associated with the activities described above. As the 27439 molecules have homology to cytochrome P450, they are expected to be involved in controlling similar disorders.

[0925] A cytochrome P450 CYP27 has been reported to play a role in the metabolism of vitamin D analogs used as drugs (Guo, Y.D. *et al.* (1993) *Proc Natl Acad Sci USA* 90(18):8668-72). In addition, a cytochrome P450 has been shown to catalyze the 25-hydroxylation of vitamin D3 (Su, P. *et al.* (1990) *DNA Cell Biol* 9(9):657-67), which plays a role in calcium homeostasis (Monkawa, T. (1997) *Biochem Biophys Res Commun* 239(2):527-33). Thus, without being bound by theory, the 27439 molecules may play a role in treating disturbances in vitamin D metabolism or diseases related to calcium metabolism such as vitamin D-dependency rickets, and/or may be important in the metabolism of vitamin D analogs used as drugs.

[0926] To identify the presence of a "Cytochrome P450 family" domain in a 27439 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters. For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.*, (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*, (1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*, (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*, (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

[0927] A 27439 polypeptide can include a "Cytochrome P450 family domain" or regions homologous with a "Cytochrome P450 family domain". As used herein, the term "Cytochrome P450 family domain" includes an amino acid sequence of about 100-500 amino acid residues in length and having a bit score for the alignment of the sequence to the Cytochrome P450 family domain (HMM) of at least 8. Preferably, a Cytochrome P450 family domain includes at least about 100-500 amino acids, more preferably about 200-500 amino acid residues, or about 300-475 amino acids and has a bit score for the alignment of the sequence to the Cytochrome P450 family domain (HMM) of at least 16, 50, 100, 200 or greater. The Cytochrome P450 family domain (HMM) has been assigned the PFAM Accession PF00067. The Cytochrome P450 family domain (amino acids 69 to 533 of SEQ

ID NO:17) of human 27439 aligns with a consensus amino acid sequence (SEQ ID NO:19) derived from a hidden Markov model.

[0928] In a preferred embodiment 27439 polypeptide or protein has a “Cytochrome P450 family domain” or a region which includes at least about 100-500 more preferably about 200-500 or 300-475 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a “Cytochrome P450 family domain,” e.g., the Cytochrome P450 family domain of human 27439 (e.g., amino acid residues 69-533 of SEQ ID NO:17).

[0929] For further identification of domains in a 27439 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against a database of domains, e.g., the ProDom database (Corpet et al. (1999), *Nucl. Acids Res.* 27:263-267). The ProDom protein domain database consists of an automatic compilation of homologous domains. Current versions of ProDom are built using recursive PSI-BLAST searches (Altschul SF *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Gouzy *et al.* (1999) 23:333-340) of the SWISS-PROT 38 and TREMBL protein databases. The database automatically generates a consensus sequence for each domain.

[0930] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD330854 (“cytochrome P450 precursor oxidoreductase monooxygenase mitochondrial heme 1.14.-.- membrane cholesterol” SEQ ID NO:20, ProDomain Release 2001.1; see ProDom database online). The “cytochrome P450 precursor oxidoreductase monooxygenase mitochondrial heme 1.14.-.- membrane cholesterol” domain (amino acids 146-313 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:20) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:20 is 33% identical over amino acids 146-313 of SEQ ID NO:17.

[0931] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD322110 (“oxidoreductase CypCCCXVA1 membrane P450 reticulum 1.14.-.- 315A1 monooxygenase cytochrome microsome” SEQ ID NO:21, ProDomain Release 2001.1; see ProDom database online). The “oxidoreductase CypCCCXVA1 membrane P450 reticulum 1.14.-.- 315A1 monooxygenase cytochrome microsome” domain (amino acids 71-407 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:21) derived from

a hidden Markov model. The consensus sequence for SEQ ID NO:21 is 24% identical over amino acids 71-407 of SEQ ID NO:17.

[0932] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD008481(“25-hydroxyvitamin sterol D3 cytochrome precursor mitochondrial 1.14.-.- alpha-hydroxylase” SEQ ID NO:22, ProDomain Release 2001.1; see ProDom database online). The “25-hydroxyvitamin sterol D3 cytochrome precursor mitochondrial 1.14.-.- alpha-hydroxylase” domain (amino acids 66-152 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:22) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:22 is 40% identical over amino acids 66-152 of SEQ ID NO:17.

[0933] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD186271(“cytochrome P450 monooxygenase oxidoreductase heme demethylase sterol 1.14.” SEQ ID NO:23, ProDomain Release 2001.1; see ProDom database online). The “cytochrome P450 monooxygenase oxidoreductase heme demethylase sterol 1.14.” domain (amino acids 389-464 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:23) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:23 is 42% identical over amino acids 389-464 of SEQ ID NO:17.

[0934] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD116294(“thromboxane-A synthase” SEQ ID NO:24, ProDomain Release 2001.1; see ProDom database online). The “thromboxane-A synthase” domain (amino acids 99-409 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:24) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:24 is 22% identical over amino acids 99-409 of SEQ ID NO:17.

[0935] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD317298(“oxidoreductase transit precursor membrane 302A1 P450 disembodied 1.14.-.- peptide monooxygenase” SEQ ID NO:25, ProDomain Release 2001.1; see Prodom database online). The “oxidoreductase transit precursor membrane 302A1 P450 disembodied 1.14.-.- peptide monooxygenase” domain (amino acids 68-196 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:25) derived from a hidden Markov model. The

consensus sequence for SEQ ID NO:25 is 29% identical over amino acids 68-196 of SEQ ID NO:17.

[0936] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD134174(“cytochrome P450 P450-DIT2 oxidoreductase monooxygenase sporulation heme” SEQ ID NO:26, ProDomain Release 2001.1; see ProDom database online). The “cytochrome P450 P450-DIT2 oxidoreductase monooxygenase sporulation heme” domain (amino acids 88-410 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:26) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:26 is 19% identical over amino acids 88-410 of SEQ ID NO:17.

[0937] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD196171(“P450 cytochrome” SEQ ID NO:27, ProDomain Release 2001.1; see ProDom database online). The “P450 cytochrome” domain (amino acids 396-537 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NO:27) derived from a hidden Markov model. The consensus sequence for SEQ ID NO:27 is 26% identical over amino acids 396-537 of SEQ ID NO:17.

[0938] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD319292(“P450 cytochrome” SEQ ID NOs:28 and 29, ProDomain Release 2001.1; see ProDom database online). The “P450 cytochrome” domain (amino acids 344-411 and 69-169 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NOs:28 and 29) derived from a hidden Markov model. The consensus sequence for SEQ ID NOs:28 and 29 are 30% and 22% identical, respectively over amino acids 344-411 and 69-169 of SEQ ID NO:17.

[0939] A BLAST search was performed against the HMM database resulting in the identification of regions homologous to ProDom family PD099144(“cytochrome P450 oxidoreductase 1.14.-.- monooxygenase heme transmembrane n-hydroxylating 79B2 Cyp79F2” SEQ ID NOs:30 and 31, ProDomain Release 2001.1; see ProDom database online). The “cytochrome P450 oxidoreductase 1.14.-.- monooxygenase heme transmembrane n-hydroxylating 79B2 Cyp79F2” domain (amino acids 153-274 and 458-466 of SEQ ID NO:17) of human 27439 aligns with consensus amino acid sequences (SEQ ID NOs:30 and 31) derived from a hidden Markov model. The consensus sequence for SEQ

ID NOs:30 and 31 are 25% and 77%, respectively identical over amino acids 153-274 and 458-466 of SEQ ID NO:17.

[0940] In one embodiment, a 27439 protein includes at least one transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues in length that spans a phospholipid membrane. More preferably, a transmembrane domain includes about at least 18, 20, 22, 24, 25, 30, 35 or 40 amino acid residues and spans a phospholipid membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an α -helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, pfam.wustl.edu/cgi-bin/getdesc?name=7tm-1, and Zagotta W.N. et al., (1996) *Annual Rev. Neurosci.* 19: 235-63, the contents of which are incorporated herein by reference.

[0941] In a preferred embodiment, a 27439 polypeptide or protein has at least one transmembrane domain or a region which includes at least 18, 20, 22, 24, 25, 30, 35 or 40 amino acid residues and has at least about 60%, 70% 80% 90% 95%, 99%, or 100% homology with a "transmembrane domain," e.g., at least one transmembrane domain of human 27439 (e.g., amino acid residues 316-332 of SEQ ID NO:17).

[0942] In another embodiment, a 27439 protein includes at least one "non-transmembrane domain." As used herein, "non-transmembrane domains" are domains that reside outside of the membrane. When referring to plasma membranes, non-transmembrane domains include extracellular domains (i.e., outside of the cell) and intracellular domains (i.e., within the cell). When referring to membrane-bound proteins found in intracellular organelles (e.g., mitochondria, endoplasmic reticulum, peroxisomes and microsomes), non-transmembrane domains include those domains of the protein that reside in the cytosol (i.e., the cytoplasm), the lumen of the organelle, or the matrix or the intermembrane space (the latter two relate specifically to mitochondria organelles). The C-terminal amino acid residue of a non-transmembrane domain is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally-occurring 27439, or 27439-like protein.

[0943] In a preferred embodiment, a 27439 polypeptide or protein has a "non-transmembrane domain" or a region which includes at least about 1-350, preferably about 50-350, more preferably about 100-350, and even more preferably about 150-350 amino acid residues, and has at least about 60%, 70% 80% 90% 95%, 99% or 100% homology

with a “non-transmembrane domain”, e.g., a non-transmembrane domain of human 27439 (e.g., residues 1-315 and 363-537 of SEQ ID NO:17). Preferably, a non-transmembrane domain is capable of catalytic activity (e.g., hydroxylation).

[0944] A non-transmembrane domain located at the N-terminus of a 27439 protein or polypeptide is referred to herein as an “N-terminal non-transmembrane domain.” As used herein, an “N-terminal non-transmembrane domain” includes an amino acid sequence having about 1-315, preferably about 100-315, more preferably about 200-315, or even more preferably about 250-315 amino acid residues in length and is located outside the boundaries of a membrane. For example, an N-terminal non-transmembrane domain is located at about amino acid residues 1-315 of SEQ ID NO:17.

[0945] Similarly, a non-transmembrane domain located at the C-terminus of a 27439 protein or polypeptide is referred to herein as a “C-terminal non-transmembrane domain.” As used herein, an “C-terminal non-transmembrane domain” includes an amino acid sequence having about 1-180, preferably about 50-180, preferably about 100-180, more preferably about 140-180 amino acid residues in length and is located outside the boundaries of a membrane. For example, an C-terminal non-transmembrane domain is located at about amino acid residues 363-537 of SEQ ID NO:17.

[0946] In a preferred embodiment, a 27439 family member can include at least one cytochrome P450 family domain (PFAM Accession Number PF00067). Furthermore, a 27439 family member can include at least one N-glycosylation site (PS00001); at least one cAMP- and cGMP-dependent protein kinase phosphorylation site (PS00004); at least one, and preferably two protein kinase C phosphorylation sites (PS00005); at least one, two, three, four, five, six, and preferably seven casein kinase II phosphorylation sites (PS00006); at least one, and preferably two tyrosine kinase phosphorylation sites (PS00007); at least one, two, three, four, five, six, seven, and preferably eight N-myristoylation sites (PS00008); at least one, and preferably two amidation sites (PS00009); and at least one cytochrome P450 cysteine heme-iron ligand signature site (PS00086).

[0947] As the 27439 polypeptides of the invention may modulate 27439-mediated activities, they may be useful for developing novel diagnostic and therapeutic agents for 27439-mediated or related disorders, as described below.

[0948] As used herein, a “Cytochrome P450 family-associated disorder” includes a disorder, disease or condition which is caused or characterized by a misregulation (e.g., downregulation or upregulation) of a Cytochrome P450 family-mediated activity.

Cytochrome P450 family-associated disorders can detrimentally affect cellular functions such as cellular proliferation, growth, differentiation, or migration, cellular regulation of homeostasis, inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; and protection of cells from toxic compounds (*e.g.*, carcinogens, toxins, mutagens, and toxic byproducts of metabolic activity (*e.g.*, reactive oxygen species)). Accordingly, 27439 protein may mediate various disorders, including cellular proliferative and/or differentiative disorders, brain disorders, heart disorders, cardiovascular disorders, bone metabolism, and pain or metabolic disorders.

[0949] Clinical toxicity may occur from a patient having a particular isoform of a Cytochrome P450 family enzyme. This is a result of improper or poor metabolism of a drug or alternatively competitive inhibition of drug metabolism may occur if two different drugs are metabolized by the same isoform. Other members of the Cytochrome P450 family have been shown to anticipate adverse drug interactions. For example, patients with poor metabolizing isoforms of Cyp2D6 has been shown to suffer cardiotoxic effects when the drug desipramine is administered. Cyp2D6 is also known to be inhibited by cimetidine, and the drug families of the selective serotonin reuptake inhibitors and tricyclic antidepressants. A patient that takes a combination of any of these inhibitors with a drug specifically metabolized by Cyp2D6 will result in an increased half-life of the drug. The 27439 polypeptides function in a similar manner and are expected to play a role in indication of adverse drug interactions in a patient with 27439 polypeptides. Thus, the 27439 compositions of the invention (*e.g.*, nucleic acids, polypeptides, proteins, antibodies) can be used to modulate drug metabolism, either *in vitro* in cell or *in vivo* in subjects, and furthermore can be used in screening assays to identify agents that modulate drug metabolism; as well as in detection or diagnostic assays for disorders involving drug metabolism.

[0950] Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

[0951] As used herein, the terms "cancer", "hyperproliferative" and "neoplastic" refer to cells having the capacity for autonomous growth, *i.e.*, an abnormal state or condition characterized by rapidly proliferating cell growth. Hyperproliferative and neoplastic disease states may be categorized as pathologic, *i.e.*, characterizing or constituting a disease state, or

may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. "Pathologic hyperproliferative" cells occur in disease states characterized by malignant tumor growth. Examples of non-pathologic hyperproliferative cells include proliferation of cells associated with wound repair.

[0952] The terms "cancer" or "neoplasms" include malignancies of the various organ systems, such as affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genitourinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus.

[0953] The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures.

[0954] The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[0955] The 27439 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of proliferative disorders. E.g., such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L., (1991) *Crit. Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-

lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[0956] Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem,

including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

[0957] Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and

myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

[0958] Examples of disorders involving the heart or "cardiovascular disorder" include, but are not limited to, a disease, disorder, or state involving the cardiovascular system, *e.g.*, the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, *e.g.*, by a thrombus. Examples of cardiovascular disorders include but are not limited to, hypertension, atherosclerosis, coronary artery spasm, coronary artery disease, arrhythmias, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus

(Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, disorders involving cardiac transplantation, and congestive heart failure.

[0959] A cardiovascular disease or disorder also includes an endothelial cell disorder.

[0960] As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, *e.g.*, proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, *e.g.*, TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory diseases (*e.g.*, rheumatoid arthritis).

[0961] Aberrant expression and/or activity of 27439 molecules may mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, *e.g.*, bone formation, bone resorption, etc., which may ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 27439 molecules effects in bone cells, *e.g.* osteoclasts and osteoblasts, that may in turn result in bone formation and degeneration. For example, 27439 molecules may support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 27439 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus may be used to treat bone disorders. Examples of such disorders include, but are not limited to,

osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[0962] Additionally, 27439 may play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L., (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

[0963] The 27439 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:17 are collectively referred to as "polypeptides or proteins of the invention" or "27439 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "27439 nucleic acids." 27439 molecules refer to 27439 nucleic acids, polypeptides, and antibodies.

[0964] As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0965] The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid

molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0966] As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred, example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:16, or SEQ ID NO:18, corresponds to a naturally-occurring nucleic acid molecule.

[0967] As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[0968] As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 27439 protein, preferably a mammalian 27439 protein, and can further include non-coding regulatory sequences, and introns.

[0969] An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 27439 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-27439 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-27439 chemicals. When the 27439 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[0970] A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 27439(e.g., the sequence of SEQ ID NO:16, SEQ ID NO:18) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the Cytochrome P450 family domain, are predicted to be particularly unamenable to alteration.

[0971] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 27439 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 27439 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 27439 biological activity to identify mutants

that retain activity. Following mutagenesis of SEQ ID NO:16, SEQ ID NO:18, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[0972] As used herein, a "biologically active portion" of a 27439 protein includes a fragment of a 27439 protein which participates in an interaction between a 27439 molecule and a non-27439 molecule. Biologically active portions of a 27439 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 27439 protein, e.g., the amino acid sequence shown in SEQ ID NO:17, which include less amino acids than the full length 27439 proteins, and exhibit at least one activity of a 27439 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 27439 protein, e.g., Cytochrome P450 family activity. A biologically active portion of a 27439 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 27439 protein can be used as targets for developing agents which modulate a 27439 mediated activity, e.g., Cytochrome P450 family activity.

[0973] Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

[0974] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 27439 amino acid sequence of SEQ ID NO:17 having 537 amino acid residues, at least 161, preferably at least 215, more preferably at least 269, even more preferably at least 322, and even more preferably at least 376, 430, 483 or 537 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions

shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0975] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0976] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0977] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al., (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 27439 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 27439 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0978] Particular 27439 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:17. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:17 are termed substantially identical.

[0979] In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:16 or 18 are termed substantially identical.

[0980] "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[0981] "Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

[0982] A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[0983] Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

[0984] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 27439 polypeptide described herein, e.g., a full length 27439 protein or a fragment thereof, e.g., a biologically active portion of 27439 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify a nucleic acid molecule encoding a polypeptide of the invention, 27439 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

[0985] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:16, or a portion of any of these nucleotide sequences. In one embodiment, the nucleic acid molecule includes sequences encoding the human 27439 protein (i.e., "the coding region", from nucleotides 1-1614 of SEQ ID NO:16, including the terminal codon). In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the mature protein of SEQ ID NO:17.

[0986] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18, thereby forming a stable duplex.

[0987] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the

nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18. In the case of an isolated nucleic acid molecule which is longer than or equivalent in length to the reference sequence, e.g., SEQ ID NO:16, or SEQ ID NO:18, the comparison is made with the full length of the reference sequence. Where the isolated nucleic acid molecule is shorter than the reference sequence, e.g., shorter than SEQ ID NO:16, or SEQ ID NO:18, the comparison is made to a segment of the reference sequence of the same length (excluding any loop required by the homology calculation).

27439 Nucleic Acid Fragments

[0988] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:16, SEQ ID NO:18. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 27439 protein, e.g., an immunogenic or biologically active portion of a 27439 protein. A fragment can comprise: nucleotides 205-1599 of SEQ ID NO:16, which encodes a Cytochrome P450 family domain of human 27439. The nucleotide sequence determined from the cloning of the 27439 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 27439 family members, or fragments thereof, as well as 27439 homologues, or fragments thereof, from other species.

[0989] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 150 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not be construed as encompassing those fragments that may have been disclosed prior to the invention.

[0990] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, the nucleic acid fragment can include a Cytochrome P450 family domain. In a preferred embodiment the fragment is at least, 50, 100, 200, 300, 400, 500, 600, 700, or 900 base pairs in length.

[0991] 27439 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:16, SEQ ID NO:18, or of a naturally occurring allelic variant or mutant of SEQ ID NO:16, SEQ ID NO:18.

[0992] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[0993] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes a Cytochrome P450 family domain (e.g., about nucleotides 205-1599 of SEQ ID NO:16).

[0994] In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 27439 sequence, e.g., a region described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a naturally occurring variant. E.g., primers suitable for amplifying all or a portion of any of the following regions are provided: a Cytochrome P450 family domain (e.g., about nucleotides 205-1599 of SEQ ID NO:16).

[0995] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[0996] A nucleic acid fragment encoding a "biologically active portion of a 27439 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:16, SEQ ID NO:18, which encodes a polypeptide having a 27439 biological activity (e.g., the biological activities of the 27439 proteins as described herein), expressing the encoded portion of the 27439 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 27439 protein. For example, a nucleic acid fragment encoding a biologically active portion of 27439 includes a Cytochrome P450 family domain (e.g., about nucleotides 205-1599 of SEQ ID NO:16). A nucleic acid

fragment encoding a biologically active portion of a 27439 polypeptide, may comprise a nucleotide sequence which is greater than 300-1200 or more nucleotides in length.

[0997] In preferred embodiments, nucleic acids include a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:16, or SEQ ID NO:18.

27439 Nucleic Acid Variants

[0998] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:16, SEQ ID NO:18. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 27439 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:17. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[0999] Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non preferred, for a particular expression system. E.g., the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in E. coli, yeast, human, insect, or CHO cells.

[01000] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

[01001] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:16, SEQ ID NO:18, e.g., as follows: by at least one but less than 10, 20, 30, or 40

nucleotides; at least one but less than 1%, 5%, 10% or 20% of the in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[01002] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the amino acid sequence shown in SEQ ID NO:17 or a fragment of this sequence. Such nucleic acid molecules can readily be obtained as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:18 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 27439 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 27439 gene. Preferred variants include those that are correlated with Cytochrome P450 family activity.

[01003] Allelic variants of 27439, e.g., human 27439, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 27439 protein within a population that maintain the ability to modulate the phosphorylation state of itself or another protein or polypeptide. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:17, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 27439, e.g., human 27439, protein within a population that do not have the ability to activate hydroxylation. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:17, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

[01004] Moreover, nucleic acid molecules encoding other 27439 family members and, thus, which have a nucleotide sequence which differs from the 27439 sequences of SEQ ID NO:16, SEQ ID NO:18 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 27439 Nucleic Acid Molecules

[01005] In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 27439. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 27439 coding strand, or to only a portion thereof (e.g., the coding region of human 27439 corresponding to SEQ ID NO:18).. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 27439 (e.g., the 5' and 3' untranslated regions).

[01006] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 27439 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 27439 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 27439 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[01007] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[01008] The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 27439 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target

selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[01009] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al., (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al., (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., (1987) *FEBS Lett.* 215:327-330).

[1000] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 27439-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 27439 cDNA disclosed herein (i.e., SEQ ID NO:16, or SEQ ID NO:18), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 27439-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, 27439 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[1001] 27439 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 27439 (e.g., the 27439 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 27439 gene in target cells. See generally, Helene, C., (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al., (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J., (1992) *Bioassays* 14(12):807-15. The potential sequences that can be targeted for triple helix formation can

be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[1002] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[1003] A 27439 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. et al., (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. et al., (1996) *supra*; Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675.

[1004] PNAs of 27439 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 27439 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B., (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. et al., (1996) *supra*; Perry-O'Keefe *supra*).

[1005] In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al., (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered

cleavage agents (See, e.g., Krol et al., (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon, (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[1006] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 27439 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 27439 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

Isolated 27439 Polypeptides

[1007] In another aspect, the invention features, an isolated 27439 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-27439 antibodies. 27439 protein can be isolated from cells or tissue sources using standard protein purification techniques. 27439 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

[1008] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same postranslational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of postranslational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

[1009] In a preferred embodiment, a 27439 polypeptide has one or more of the following characteristics: i) activation of an enzymatic hydroxylation activity; ii) it has a molecular weight, e.g., a deduced molecular weight, amino acid composition or other physical characteristic of the polypeptide of SEQ ID NO:17; iii) it has an overall sequence similarity of at least 50%, preferably at least 60%, more preferably at least 70, 80, 90, or 95%, with a polypeptide of SEQ ID NO:17; iv) it has a Cytochrome P450 family domain which preferably has an overall sequence similarity of about 70%, 80%, 90% or 95% with

amino acid residues 8-231 of SEQ ID NO:17; v) it has at least 70%, preferably 80%, and vi) most preferably 95% of the cysteines found in the amino acid sequence of the native protein.

[1010] In a preferred embodiment the 27439 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:17. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:17 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:17. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the Cytochrome P450 family domain. In another embodiment one or more differences are in the Cytochrome P450 family domain.

[1011] Other embodiments include a protein that contain one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity. Such 27439 proteins differ in amino acid sequence from SEQ ID NO:17, yet retain biological activity.

[1012] In one embodiment, a biologically active portion of a 27439 protein includes a Cytochrome P450 family domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 27439 protein.

[1013] In a preferred embodiment, the 27439 protein has an amino acid sequence shown in SEQ ID NO:17. In other embodiments, the 27439 protein is substantially identical to SEQ ID NO:17. In yet another embodiment, the 27439 protein is substantially identical to SEQ ID NO:17 and retains the functional activity of the protein of SEQ ID NO:17, as described in detail above. Accordingly, in another embodiment, the 27439 protein is a protein which includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:17.

27439 Chimeric or Fusion Proteins

[1014] In another aspect, the invention provides 27439 chimeric or fusion proteins. As used herein, a 27439 "chimeric protein" or "fusion protein" includes a 27439 polypeptide linked to a non-27439 polypeptide. A "non-27439 polypeptide" refers to a polypeptide

having an amino acid sequence corresponding to a protein which is not substantially homologous to the 27439 protein, e.g., a protein which is different from the 27439 protein and which is derived from the same or a different organism. The 27439 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 27439 amino acid sequence. In a preferred embodiment, a 27439 fusion protein includes at least one (or two) biologically active portion of a 27439 protein. The non-27439 polypeptide can be fused to the N-terminus or C-terminus of the 27439 polypeptide.

[1015] The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-27439 fusion protein in which the 27439 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 27439. Alternatively, the fusion protein can be a 27439 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 27439 can be increased through use of a heterologous signal sequence.

[1016] Fusion proteins can include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin.

[1017] The 27439 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 27439 fusion proteins can be used to affect the bioavailability of a 27439 substrate. 27439 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 27439 protein; (ii) mis-regulation of the 27439 gene; and (iii) aberrant post-translational modification of a 27439 protein.

[1018] Moreover, the 27439-fusion proteins of the invention can be used as immunogens to produce anti-27439 antibodies in a subject, to purify 27439 ligands and in screening assays to identify molecules which inhibit the interaction of 27439 with a 27439 substrate.

[1019] Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 27439-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 27439 protein.

Variants of 27439 Proteins

[1020] In another aspect, the invention also features a variant of a 27439 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of

the 27439 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 27439 protein. An agonist of the 27439 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 27439 protein. An antagonist of a 27439 protein can inhibit one or more of the activities of the naturally occurring form of the 27439 protein by, for example, competitively modulating a 27439-mediated activity of a 27439 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 27439 protein.

[1021] Variants of a 27439 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 27439 protein for agonist or antagonist activity.

[1022] Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 27439 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 27439 protein.

[1023] Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[1024] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 27439 variants (Arkin and Yourvan, (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al., (1993) *Protein Engineering* 6(3):327-331).

[1025] Cell based assays can be exploited to analyze a variegated 27439 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 27439 in a substrate-dependent manner. The transfected cells are then contacted with 27439 and the effect of the expression of the mutant on signaling by the 27439 substrate can be detected, e.g., by measuring Cytochrome P450 family activity. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 27439 substrate, and the individual clones further characterized.

[1026] In another aspect, the invention features a method of making a 27439 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 27439 polypeptide, e.g., a naturally occurring 27439 polypeptide. The method includes: altering the sequence of a 27439 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[1027] In another aspect, the invention features a method of making a fragment or analog of a 27439 polypeptide a biological activity of a naturally occurring 27439 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 27439 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-27439 Antibodies

[1028] In another aspect, the invention provides an anti-27439 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin.

[1029] The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

[1030] A full-length 27439 protein or, antigenic peptide fragment of 27439 can be used as an immunogen or can be used to identify anti-27439 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 27439 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:17 and encompasses an epitope of 27439. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[1031] Fragments of 27439 which include, e.g., residues 440-460 of SEQ ID NO:17 can be, e.g., used as immunogens, or used to be hydrophilic regions of the 27439 protein. Similarly, a fragment of 27439 which includes, e.g., residues 318-328 of SEQ ID NO:17 can be used to make an antibody against what is believed to be a hydrophobic region of the 27439 protein; a fragment of 27439 which includes, e.g., residues 69-533 of SEQ ID NO:17 can be used to make an antibody against what is believed to be the Cytochrome P450 family region of the 27439 protein.

[1032] Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

[1033] In a preferred embodiment the antibody fails to bind an Fc receptor, e.g. it is a type which does not support Fc receptor binding or has been modified, e.g., by deletion or other mutation, such that it does not have a functional Fc receptor binding region.

[1034] Preferred epitopes encompassed by the antigenic peptide are regions of 27439 are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 27439 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 27439 protein and are thus likely to constitute surface residues useful for targeting antibody production.

[1035] In a preferred embodiment the antibody binds an epitope on any domain or region on 27439 proteins described herein.

[1036] Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients.

[1037] Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.*

(1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

[1038] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. *See*, for example, Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93; and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

[1039] Completely human antibodies that recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described by Jespers *et al.* (1994) *Bio/Technology* 12:899-903).

[1040] The anti-27439 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered (see, for example, Colcher, D. *et al.* (1999) *Ann. N Y Acad. Sci.* 880:263-80; and Reiter, Y. (1996) *Clin. Cancer Res.* 2:245-52). The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 27439 protein.

[1041] In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it is an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, *e.g.*, it has a mutagenized or deleted Fc receptor binding region. An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy

anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

[1042] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[1043] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[1044] An anti-27439 antibody (*e.g.*, monoclonal antibody) can be used to isolate 27439 by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-27439 antibody can be used to detect 27439 protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-27439 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance (*i.e.*, antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase;

examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

[1045] In preferred embodiments, an antibody can be made by immunizing with a purified 27439 antigen, or a fragment thereof, *e.g.*, a fragment described herein, a membrane associated antigen, tissues, *e.g.*, crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[1046] Antibodies which bind only a native 27439 protein, only denatured or otherwise non-native 27439 protein, or which bind both, are within the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes sometimes can be identified by identifying antibodies which bind to native but not denatured 27439 protein.

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[1047] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, *e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses.

[1048] A vector can include a 27439 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby

produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 27439 proteins, mutant forms of 27439 proteins, fusion proteins, and the like).

[1049] The recombinant expression vectors of the invention can be designed for expression of 27439 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[1050] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S., (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[1051] Purified fusion proteins can be used in 27439 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 27439 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[1052] To maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinant

protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[1053] The 27439 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

[1054] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[1055] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton, (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al., (1983) *Cell* 33:729-740; Queen and Baltimore, (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al., (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166).

Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss, (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman, (1989) *Genes Dev.* 3:537-546).

[1056] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or

attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

[1057] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 27439 nucleic acid molecule within a recombinant expression vector or a 27439 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[1058] A host cell can be any prokaryotic or eukaryotic cell. For example, a 27439 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[1059] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation

[1060] A host cell of the invention can be used to produce (i.e., express) a 27439 protein. Accordingly, the invention further provides methods for producing a 27439 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 27439 protein has been introduced) in a suitable medium such that a 27439 protein is produced. In another embodiment, the method further includes isolating a 27439 protein from the medium or the host cell.

[1061] In another aspect, the invention features, a cell or purified preparation of cells which include a 27439 transgene, or which otherwise misexpress 27439. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 27439

transgene, e.g., a heterologous form of a 27439, e.g., a gene derived from humans (in the case of a non-human cell). The 27439 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous 27439, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed 27439 alleles or for use in drug screening.

[1062] In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 27439 polypeptide.

[1063] Also provided are cells or a purified preparation thereof, e.g., human cells, in which an endogenous 27439 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 27439 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 27439 gene. For example, an endogenous 27439 gene, e.g., a gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[1064] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 27439 protein and for identifying and/or evaluating modulators of 27439 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other

transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 27439 gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

[1065] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 27439 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 27439 transgene in its genome and/or expression of 27439 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 27439 protein can further be bred to other transgenic animals carrying other transgenes.

[1066] 27439 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[1067] The invention also includes a population of cells from a transgenic animal, as discussed herein.

Uses

[1068] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

[1069] The isolated nucleic acid molecules of the invention can be used, for example, to express a 27439 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 27439 mRNA (e.g., in a biological sample) or a genetic alteration in a 27439 gene, and to modulate 27439 activity, as described further below. The 27439 proteins can be used to treat disorders characterized by insufficient or excessive production of a 27439 substrate or production of 27439 inhibitors. In addition, the

27439 proteins can be used to screen for naturally occurring 27439 substrates, to screen for drugs or compounds which modulate 27439 activity, as well as to treat disorders characterized by insufficient or excessive production of 27439 protein or production of 27439 protein forms which have decreased, aberrant or unwanted activity compared to 27439 wild-type protein. Such disorders include those characterized by aberrant signaling or aberrant, e.g., hydroxylase activity. Moreover, the anti-27439 antibodies of the invention can be used to detect and isolate 27439 proteins, regulate the bioavailability of 27439 proteins, and modulate 27439 activity.

[1070] A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 27439 polypeptide is provided. The method includes: contacting the compound with the subject 27439 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 27439 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 27439 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 27439 polypeptide. Screening methods are discussed in more detail below.

Screening Assays:

[1071] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 27439 proteins, have a stimulatory or inhibitory effect on, for example, 27439 expression or 27439 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 27439 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 27439 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[1072] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 27439 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 27439 protein or polypeptide or a biologically active portion thereof.

[1073] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries [libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive] (see, e.g., Zuckermann, R.N. et al., *J. Med. Chem.* 1994, 37: 2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[1074] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.*, (1994). *J. Med. Chem.* 37:2678; Cho *et al.*, (1993) *Science* 261:1303; Carrell *et al.*, (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.*, (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.*, (1994) *J. Med. Chem.* 37:1233.

[1075] Libraries of compounds may be presented in solution (e.g., Houghten, (1992) *Biotechniques* 13:412-421), or on beads (Lam, (1991) *Nature* 354:82-84), chips (Fodor, (1993) *Nature* 364:555-556), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull *et al.*, (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith, (1990) *Science* 249:386-390); (Devlin, (1990) *Science* 249:404-406); (Cwirla *et al.*, (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici, (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

[1076] In one embodiment, an assay is a cell-based assay in which a cell which expresses a 27439 protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate 27439 activity is determined. Determining the ability of the test compound to modulate 27439 activity can be accomplished by monitoring, for example, Cytochrome P450 family activity. The cell, for example, can be of mammalian origin, e.g., human. Cell homogenates, or fractions, preferably membrane containing fractions, can also be tested.

[1077] The ability of the test compound to modulate 27439 binding to a compound, e.g., a 27439 substrate, or to bind to 27439 can also be evaluated. This can be

accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 27439 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 27439 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 27439 binding to a 27439 substrate in a complex. For example, compounds (e.g., 27439 substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[1078] The ability of a compound (e.g., a 27439 substrate) to interact with 27439 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 27439 without the labeling of either the compound or the 27439. McConnell, H. M. et al., (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 27439.

[1079] In yet another embodiment, a cell-free assay is provided in which a 27439 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 27439 protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the 27439 proteins to be used in assays of the present invention include fragments which participate in interactions with non-27439 molecules, e.g., fragments with high surface probability scores.

[1080] Soluble and/or membrane-bound forms of isolated proteins (e.g., 27439 proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane

sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[1081] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[1082] In one embodiment, assays are performed where the ability of an agent to block Cytochrome P450 family activity within a cell is evaluated.

[1083] The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, et al., U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[1084] In another embodiment, determining the ability of the 27439 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C., (1991) *Anal. Chem.* 63:2338-2345 and Szabo et al., (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[1085] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid

phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[1086] It may be desirable to immobilize either 27439, an anti-27439 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 27439 protein, or interaction of a 27439 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/27439 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 27439 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 27439 binding or activity determined using standard techniques.

[1087] Other techniques for immobilizing either a 27439 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 27439 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[1088] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously

non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[1089] In one embodiment, this assay is performed utilizing antibodies reactive with 27439 protein or target molecules but which do not interfere with binding of the 27439 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 27439 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 27439 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 27439 protein or target molecule.

[1090] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., *Trends Biochem Sci* 1993 Aug;18(8):284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. et al., eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. et al., eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., *J Mol. Recognit.* 1998 Winter;11(1-6):141-8; Hage, D.S., and Tweed, S.A., *J. Chromatogr. B Biomed. Sci. Appl.* 1997 Oct 10;699(1-2):499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[1091] In a preferred embodiment, the assay includes contacting the 27439 protein or biologically active portion thereof with a known compound which binds 27439 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 27439 protein, wherein determining the ability of the test compound to interact with a 27439 protein includes determining the ability of the test compound to preferentially bind to 27439 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

[1092] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 27439 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 27439 protein through modulation of the activity of a downstream effector of a 27439 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[1093] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[1094] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end

of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[1095] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[1096] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[1097] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[1098] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[1099] In yet another aspect, the 27439 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al., (1993) *Cell* 72:223-232; Madura et al., (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al., (1993) *Biotechniques* 14:920-924; Iwabuchi et al., (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 27439 ("27439-binding proteins" or "27439-bp") and are involved in 27439 activity. Such 27439-bps can be activators or inhibitors of signals by the 27439 proteins or 27439 targets as, for example, downstream elements of a 27439-mediated signaling pathway.

[1100] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 27439 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: 27439 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 27439-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 27439 protein.

[1101] In another embodiment, modulators of 27439 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 27439 mRNA or protein evaluated relative to the level of expression of 27439 mRNA or protein in the absence of the candidate compound. When expression of 27439 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 27439 mRNA or protein expression. Alternatively, when expression of 27439 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 27439 mRNA or protein expression. The level of 27439 mRNA or protein expression can be determined by methods described herein for detecting 27439 mRNA or protein.

[1102] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 27439 protein can be confirmed *in vivo*, e.g., in an animal.

[1103] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 27439 modulating agent, an antisense 27439 nucleic acid molecule, a 27439-specific antibody, or a 27439-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

[1104] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 27439 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[1105] The 27439 nucleotide sequences or portions thereof can be used to map the location of the 27439 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 27439 sequences with genes associated with disease.

[1106] Briefly, 27439 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 27439 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 27439 sequences will yield an amplified fragment.

[1107] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al., (1983) *Science* 220:919-924).

[1108] Other mapping strategies e.g., in situ hybridization (described in Fan, Y. et al., (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 27439 to a chromosomal location.

[1109] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

[1110] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[1111] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map

data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al., (1987) *Nature*, 325:783-787.

[1112] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 27439 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[1113] 27439 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[1114] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 27439 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[1115] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an

individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:16 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:18 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[1116] If a panel of reagents from 27439 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 27439 Sequences in Forensic Biology

[1117] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[1118] The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:16 (e.g., fragments derived from the noncoding regions of SEQ ID NO:16 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[1119] The 27439 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., a tissue containing Cytochrome P450 family activity. This can be very useful in cases where a

forensic pathologist is presented with a tissue of unknown origin. Panels of such 27439 probes can be used to identify tissue by species and/or by organ type.

[1120] In a similar fashion, these reagents, e.g., 27439 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

[1121] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[1122] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 27439.

[1123] Such disorders include, e.g., a disorder associated with the misexpression of 27439, or hydroxylase activity related disorder.

[1124] The method includes one or more of the following: i) detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 27439 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region; ii) detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 27439 gene; iii) detecting, in a tissue of the subject, the misexpression of the 27439 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA; and iv) detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 27439 polypeptide.

[1125] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 27439 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

[1126] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:16 naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 27439 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ*

hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[1127] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 27439 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 27439.

[1128] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[1129] In preferred embodiments the method includes determining the structure of a 27439 gene, an abnormal structure being indicative of risk for the disorder.

[1130] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 27439 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

[1131] The presence, level, or absence of 27439 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 27439 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 27439 protein such that the presence of 27439 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 27439 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 27439 genes; measuring the amount of protein encoded by the 27439 genes; or measuring the activity of the protein encoded by the 27439 genes.

[1132] The level of mRNA corresponding to the 27439 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[1133] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 27439 nucleic acid, such as the nucleic acid of SEQ ID NO:16, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 27439 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

[1134] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 27439 genes.

[1135] The level of mRNA in a sample that is encoded by one of 27439 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis, 1987, U.S. Patent No. 4,683,202), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189-193),

self sustained sequence replication (Guatelli et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi et al., U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[1136] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 27439 gene being analyzed.

[1137] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 27439 mRNA, or genomic DNA, and comparing the presence of 27439 mRNA or genomic DNA in the control sample with the presence of 27439 mRNA or genomic DNA in the test sample.

[1138] A variety of methods can be used to determine the level of protein encoded by 27439. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[1139] The detection methods can be used to detect 27439 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 27439 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western

blot analysis. *In vivo* techniques for detection of 27439 protein include introducing into a subject a labeled anti-27439 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[1140] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 27439 protein, and comparing the presence of 27439 protein in the control sample with the presence of 27439 protein in the test sample.

[1141] The invention also includes kits for detecting the presence of 27439 in a biological sample. For example, the kit can include a compound or agent capable of detecting 27439 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 27439 protein or nucleic acid.

[1142] For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[1143] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein-stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[1144] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 27439 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

[1145] In one embodiment, a disease or disorder associated with aberrant or unwanted 27439 expression or activity is identified. A test sample is obtained from a subject and 27439 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 27439 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 27439 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

[1146] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 27439 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a hydroxylase activity related disorder.

[1147] The methods of the invention can also be used to detect genetic alterations in a 27439 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 27439 protein activity or nucleic acid expression, such as a hydroxylase activity related disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 27439-protein, or the mis-expression of the 27439 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 27439 gene; 2) an addition of one or more nucleotides to a 27439 gene; 3) a substitution of one or more nucleotides of a 27439 gene, 4) a chromosomal rearrangement of a 27439 gene; 5) an alteration in the level of a messenger RNA transcript of a 27439 gene, 6) aberrant modification of a 27439 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 27439 gene, 8) a non-wild type level of a 27439-protein, 9) allelic loss of a 27439 gene, and 10) inappropriate post-translational modification of a 27439-protein.

[1148] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the

27439-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 27439 gene under conditions such that hybridization and amplification of the 27439-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

[1149] Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. et al., (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al., (1988) *Bio-Technology* 6:1197), or other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

[1150] In another embodiment, mutations in a 27439 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[1151] In other embodiments, genetic mutations in 27439 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two-dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. et al., (1996) *Human Mutation* 7: 244-255; Kozal, M.J. et al., (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in 27439 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes.

This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[1152] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 27439 gene and detect mutations by comparing the sequence of the sample 27439 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

[1153] Other methods for detecting mutations in the 27439 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al., (1985) *Science* 230:1242; Cotton et al., (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al., (1992) *Methods Enzymol.* 217:286-295).

[1154] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 27439 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al., (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[1155] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 27439 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al., (1989) *Proc. Natl. Acad. Sci. USA*: 86:2766, see also Cotton, (1993) *Mutat. Res.* 285:125-144; and Hayashi, (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 27439 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in

sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al., (1991) *Trends Genet.* 7:5).

[1156] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al., (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner, (1987) *Biophys. Chem.* 265:12753).

[1157] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki et al., (1986) *Nature* 324:163); Saiki et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:6230).

[1158] Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al., (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al., (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany, (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[1159] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose

patients exhibiting symptoms or family history of a disease or illness involving a 27439 gene.

Use of 27439 Molecules as Surrogate Markers

[1160] The 27439 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 27439 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 27439 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a “surrogate marker” is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[1161] The 27439 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a “pharmacodynamic marker” is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be

monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 27439 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-27439 antibodies may be employed in an immune-based detection system for a 27439 protein marker, or 27439-specific radiolabeled probes may be used to detect a 27439 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[1162] The 27439 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a “pharmacogenomic marker” is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 27439 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 27439 DNA may correlate 27439 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

[1163] The nucleic acid and polypeptides, fragments thereof, as well as anti-27439 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[1164] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[1165] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by

the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[1166] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[1167] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[1168] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[1169] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[1170] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[1171] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[1172] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[1173] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that

targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1174] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[1175] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[1176] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is

described by Cruikshank et al., ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[1177] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[1178] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[1179] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, teniposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine,

lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[1180] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[1181] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[1182] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al., (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[1183] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment:

[1184] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 27439 expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 27439 molecules of the present invention or 27439 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

[1185] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 27439 expression or activity, by administering to the subject a 27439 or an agent which modulates 27439 expression or at least one 27439 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 27439 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 27439 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 27439 aberrance, for example, a 27439, 27439 agonist or 27439 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[1186] It is possible that some 27439 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[1187] As discussed, successful treatment of 27439 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 27439 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and FAb, F(ab')₂ and FAb expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[1188] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[1189] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[1190] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 27439 expression is through the use of aptamer molecules specific for 27439 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, et al., *Curr. Opin. Chem. Biol.* 1997, 1(1): 5-9; and Patel, D.J., *Curr. Opin. Chem. Biol.* 1997

Jun;1(1):32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 27439 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[1191] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 27439 disorders. For a description of antibodies, see the Antibody section above.

[1192] In circumstances wherein injection of an animal or a human subject with a 27439 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 27439 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D., *Ann. Med.* 1999;31(1):66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A., *Cancer Treat. Res.* 1998;94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 27439 protein. Vaccines directed to a disease characterized by 27439 expression may also be generated in this fashion.

[1193] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco et al., (1993, *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[1194] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 27439 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

[1195] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for

determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1196] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[1197] Another example of determination of effective dose for an individual is the ability to directly assay levels of “free” and “bound” compound in the serum of the test subject. Such assays may utilize antibody mimics and/or “biosensors” that have been created through molecular imprinting techniques. The compound which is able to modulate 27439 activity is used as a template, or “imprinting molecule”, to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated “negative image” of the compound and is able to selectively rebinding the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. et al., (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J., (1994) *Trends in Polymer Science* 2:166-173. Such “imprinted” affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. et al., (1993) *Nature* 361:645-647. Through the use of isotope-labeling,

the “free” concentration of compound which modulates the expression or activity of 27439 can be readily monitored and used in calculations of IC_{50} .

[1198] Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . A rudimentary example of such a “biosensor” is discussed in Kriz, D. et al., (1995) *Analytical Chemistry* 67:2142-2144.

[1199] Another aspect of the invention pertains to methods of modulating 27439 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 27439 or agent that modulates one or more of the activities of 27439 protein activity associated with the cell. An agent that modulates 27439 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 27439 protein (e.g., a 27439 substrate or receptor), a 27439 antibody, a 27439 agonist or antagonist, a peptidomimetic of a 27439 agonist or antagonist, or other small molecule.

[1200] In one embodiment, the agent stimulates one or more 27439 activities. Examples of such stimulatory agents include active 27439 protein and a nucleic acid molecule encoding 27439. In another embodiment, the agent inhibits one or more 27439 activities. Examples of such inhibitory agents include antisense 27439 nucleic acid molecules, anti-27439 antibodies, and 27439 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 27439 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 27439 expression or activity. In another embodiment, the method involves administering a 27439 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 27439 expression or activity.

[1201] Stimulation of 27439 activity is desirable in situations in which 27439 is abnormally downregulated and/or in which increased 27439 activity is likely to have a beneficial effect. For example, stimulation of 27439 activity is desirable in situations in

which a 27439 is downregulated and/or in which increased 27439 activity is likely to have a beneficial effect. Likewise, inhibition of 27439 activity is desirable in situations in which 27439 is abnormally upregulated and/or in which decreased 27439 activity is likely to have a beneficial effect.

[1202] The 27439 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, brain disorders, heart disorders, cardiovascular disorders, bone metabolism, and pain or metabolic disorders, as described above, as well as disorders associated with hormonal disorders, immune disorders, hematopoietic disorders, liver disorders, blood vessel disorders, viral diseases, and platelet disorders.

[1203] Cytochrome P450 family-associated or related disorders also include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (*e.g.*, growth disorders), thyroid disorders (*e.g.*, hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (*e.g.*, disorders which affect the organs of the reproductive system, *e.g.*, the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject; disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

[1204] Cytochrome P450 family-associated or related disorders also include immune disorders, such as autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency.

[1205] Examples of hematopoietic disorders include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, Crohn's disease, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma, allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute

necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[1206] Disorders which may be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein may be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyl dopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[1207] Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as

hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

[1208] Additionally, 27439 molecules may play an important role in the etiology of certain viral diseases, including but not limited to, Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 27439 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 27439 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

Pharmacogenomics

[1209] The 27439 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 27439 activity (e.g., 27439 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 27439 associated disorders (e.g., hydroxylase activity related disorders) associated with aberrant or unwanted 27439 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe

toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 27439 molecule or 27439 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 27439 molecule or 27439 modulator.

[1210] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. et al. (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11) :983-985 and Linder, M.W. et al. (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[1211] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high-resolution map can be generated from a combination of some ten million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be

tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[1212] Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 27439 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[1213] Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 27439 molecule or 27439 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[1214] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 27439 molecule or 27439 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[1215] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 27439 genes of the present invention, wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 27439 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., cancer cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

[1216] Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 27439 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 27439 gene expression, protein levels, or upregulate 27439 activity, can be monitored in clinical trials of

subjects exhibiting decreased 27439 gene expression, protein levels, or downregulated 27439 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 27439 gene expression, protein levels, or downregulate 27439 activity, can be monitored in clinical trials of subjects exhibiting increased 27439 gene expression, protein levels, or upregulated 27439 activity. In such clinical trials, the expression or activity of a 27439 gene, and preferably, other genes that have been implicated in, for example, a 27439-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

[1217] In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which express or mis express 27439 or from a cell or subject in which a 27439 mediated response has been elicited, e.g., by contact of the cell with 27439 nucleic acid or protein, or administration to the cell or subject 27439 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 27439 nucleic acid, polypeptide, or antibody); providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 27439 (or does not express as highly as in the case of the 27439 positive plurality of capture probes) or from a cell or subject which in which a 27439 mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 27439 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

[1218] The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

[1219] The method can include contacting the 27439 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

[1220] The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 27439. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder. 27439 is associated with Cytochrome P450 family activity, thus it is useful for disorders associated with abnormal Cytochrome P450 family activity.

[1221] The method can be used to detect SNPs, as described above.

[1222] In another aspect, the invention features, a method of analyzing 27439, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 27439 nucleic acid or amino acid sequence; comparing the 27439 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 27439.

[1223] The method can include evaluating the sequence identity between a 27439 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet. Preferred databases include GenBank™ and SwissProt.

[1224] In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 27439. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

[1225] The sequences of 27439 molecules are provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 27439 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, *e.g.*, an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exist in nature or in purified form.

[1226] A 27439 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having thereon 27439 sequence information of the present invention.

[1227] As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus of other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

[1228] As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 27439 sequence information.

[1229] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a 27439 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-

available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[1230] By providing the 27439 nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[1231] The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, wherein the method comprises the steps of determining 27439 sequence information associated with the subject and based on the 27439 sequence information, determining whether the subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1232] The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a disease associated with 27439, wherein the method comprises the steps of determining 27439 sequence information associated with the subject, and based on the 27439 sequence information, determining whether the subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

[1233] The present invention also provides in a network, a method for determining whether a subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, said method comprising the steps of receiving 27439 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 27439 and/or corresponding to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, and based on one or more of the phenotypic information, the 27439 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1234] The present invention also provides a business method for determining whether a subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, said method comprising the steps of receiving information related to 27439 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 27439 and/or related to a Cytochrome P450 family-associated or another 27439-associated disease or disorder, and based on one or more of the phenotypic information, the 27439 information, and the acquired information, determining whether the subject has a Cytochrome P450 family-associated or another 27439-associated disease or disorder or a pre-disposition to a Cytochrome P450 family-associated or another 27439-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1235] The invention also includes an array comprising a 27439 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 27439. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

[1236] In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[1237] In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a Cytochrome P450 family-associated or another 27439-associated disease or disorder, progression of Cytochrome P450 family-associated or another 27439-associated disease or disorder, and processes, such as a cellular transformation associated with the Cytochrome P450 family-associated or another 27439-associated disease or disorder.

[1238] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of 27439 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[1239] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 27439) that could serve as a molecular target for diagnosis or therapeutic intervention.

[1240] As used herein, a “target sequence” can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[1241] Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

[1242] Thus, the invention features a method of making a computer readable record of a sequence of a 27439 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[1243] In another aspect, the invention features a method of analyzing a sequence. The method includes: providing a 27439 sequence, or record, in computer readable form; comparing a second sequence to the 27439 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, *e.g.*, determining if the 27439 sequence includes a sequence being compared. In a preferred embodiment the 27439 or second sequence is stored on a first computer, *e.g.*, at a first site and the comparison is performed, read, or recorded on a second computer, *e.g.*, at a second site. *E.g.*, the 27439 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the

transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[1244] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

Example 1: Identification and Characterization of Human 27439 cDNAs

[1245] The human 27439 sequence (SEQ ID NO:16), which is approximately 1976 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1614 nucleotides (nucleotides 1-1614 of SEQ ID NO:16; SEQ ID NO:18), including the terminal codon. The coding sequence encodes a 537 amino acid protein (SEQ ID NO:17).

Example 2: Tissue Distribution of 27439 mRNA

[1246] Northern blot hybridizations with various RNA samples can be performed under standard conditions and washed under stringent conditions, i.e., 0.2xSSC at 65°C. A DNA probe corresponding to all or a portion of the 27439 cDNA (SEQ ID NO:16) can be used. The DNA was radioactively labeled with ³²P-dCTP using the Prime-It Kit (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA from mouse hematopoietic and endocrine tissues, and cancer cell lines (Clontech, Palo Alto, CA) can be probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Example 3: Gene Expression Analysis

[1247] Total RNA was prepared from various human tissues by a single step extraction method using RNA STAT-60 according to the manufacturer's instructions (TelTest, Inc). Each RNA preparation was treated with DNase I (Ambion) at 37°C for 1 hour. DNase I treatment was determined to be complete if the sample required at least 38 PCR amplification cycles to reach a threshold level of fluorescence using β -2 microglobulin as an internal amplicon reference. The integrity of the RNA samples following DNase I treatment was confirmed by agarose gel electrophoresis and ethidium bromide staining.

After phenol extraction cDNA was prepared from the sample using the SUPERScript™ Choice System following the manufacturer's instructions (GibcoBRL). A negative control of RNA without reverse transcriptase was mock reverse transcribed for each RNA sample.

[1248] Human 27439 expression was measured by TaqMan® quantitative PCR (Perkin Elmer Applied Biosystems) in cDNA prepared from a variety of normal and diseased (e.g., cancerous) human tissues or cell lines.

[1249] Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the human 27439 gene. Each human 27439 gene probe was labeled using FAM (6-carboxyfluorescein), and the β 2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β 2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe for β -2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). The thermal cycler conditions were as follows: hold for 2 min at 50°C and 10 min at 95°C, followed by two-step PCR for 40 cycles of 95°C for 15 sec followed by 60°C for 1 min.

[1250] The following method was used to quantitatively calculate human 27439 gene expression in the various tissues relative to β -2 microglobulin expression in the same tissue. The threshold cycle (Ct) value is defined as the cycle at which a statistically significant increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the human 27439 gene is normalized by subtracting the Ct value of the β -2 microglobulin gene to obtain a Δ Ct value using the following formula: Δ Ct = Ct_{human 59914 and 59921} - Ct _{β -2 microglobulin}. Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the human 27439 gene. The Δ Ct value for the calibrator sample is then subtracted from Δ Ct for each tissue sample according to the following formula: $\Delta\Delta$ Ct = Δ Ct_{sample} - Δ Ct_{calibrator}. Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta$ Ct}. Expression of the target human 27439 gene in each of the tissues tested is then graphically represented as discussed in more detail below.

[1251] TaqMan real-time quantitative RT-PCR is used to detect the presence of RNA transcript corresponding to human 27439 relative to a no template control in a panel of human tissues or cells. It is found that the highest expression of 27439 orthologs are expressed in normal skin tissue, as shown in table 2. Relatively high expression is also seen in fibroblasts, osteoblasts, and hypothalamus tissue. It is also of note that there is increased expression of 27439 in breast tumor, and prostate tumor compared to normal breast and prostate tissue.

Table 2

Tissue MEAN	MEAN	β2M803	ΔCt	Expression
Aorta/Normal	39.15	20.87	18.47	0.00
Fetal Heart/Normal	32.09	19.46	12.68	0.15
Heart/Normal	40.00	20.42	19.63	0.00
Heart/CHF	39.46	19.43	20.14	0.00
Vein/Normal	40.00	22.23	17.80	0.00
SMC/Aortic	40.00	24.04	16.03	0.01
Nerve/Normal	39.59	23.39	16.35	0.01
Spinal Cord/Normal	35.05	20.50	14.51	0.04
Brain Cortex/Normal	40.00	25.28	14.86	0.03
Brain Hypothalamus/Normal	31.17	19.63	11.36	0.38
Glial Cells (Astrocytes)	37.50	21.00	16.50	0.01
Glioblastoma	33.62	16.98	16.38	0.01
Breast/Normal	34.69	20.27	14.34	0.05
Breast/Tumor	29.09	16.30	12.88	0.13
Ovary/Normal	37.95	18.70	19.27	0.00
Ovary/Tumor	34.86	18.07	16.85	0.01
Pancreas/Normal	37.87	18.08	19.96	0.00
Prostate/Normal	34.28	18.05	16.11	0.01
Prostate/Tumor	32.71	18.01	14.60	0.04
Colon/Normal	35.66	19.15	16.32	0.01
Colon/Tumor	37.45	17.23	20.16	0.00
Colon/IBD	37.25	16.42	20.77	0.00
Kidney/Normal	37.36	18.63	18.69	0.00
Liver/Normal	38.36	18.56	20.43	0.00
Liver/Fibrosis	38.14	19.32	18.85	0.00
Fetal Liver/Normal	35.71	19.96	15.89	0.02
Lung/Normal	38.57	17.24	21.20	0.00
Lung/COPD	38.69	17.62	21.07	0.00
Spleen/Normal	39.54	19.18	20.48	0.00
Tonsil/Normal	33.50	16.72	16.80	0.01
Lymph Node/Normal	34.85	17.51	17.10	0.01
Thymus/Normal	34.93	19.32	15.73	0.02
Epithelial Cells (Prostate)	35.03	22.79	12.28	0.20

Endothelial Cells (Aortic)	40.00	21.25	18.69	0.00
Skeletal Muscle/Normal	34.39	17.51	16.91	0.01
Fibroblasts (Dermal)	33.63	23.50	10.25	0.82
Skin/Normal	39.93	31.12	7.93	4.11
Adipose/Normal	39.20	21.82	17.16	0.01
Osteoblasts (Primary)	31.08	20.19	10.86	0.54
Osteoblasts (Undiff)	37.71	18.09	19.69	0.00
Osteoblasts (Diff)	36.74	17.37	19.27	0.00
Osteoclasts	31.52	15.92	15.52	0.02
NTC	40.00	40.00	0.00	0.00

Example 4: Recombinant Expression of 27439 in Bacterial Cells

[1252] In this example, 27439 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, 27439 is fused to GST and this fusion polypeptide is expressed in *E. coli*, e.g., strain PEB199. Expression of the GST-27439 fusion protein in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

Example 5: Expression of Recombinant 27439 Protein in COS Cells

[1253] To express the 27439 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire 27439 protein and an HA tag (Wilson et al. (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant protein under the control of the CMV promoter.

[1254] To construct the plasmid, the 27439 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the 27439 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the 27439 coding sequence. The PCR amplified fragment and the pcDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using

the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the 27439 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5 α , SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

[1255] COS cells are subsequently transfected with the 27439-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T.

Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the 27439 polypeptide is detected by radiolabelling (^{35}S -methionine or ^{35}S -cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labeled for 8 hours with ^{35}S -methionine (or ^{35}S -cysteine). The culture media are then collected and the cells are lysed using detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

[1256] Alternatively, DNA containing the 27439 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the 27439 polypeptide is detected by radiolabelling and immunoprecipitation using a 27439 specific monoclonal antibody.

Equivalents

[1257] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

V. 68730 AND 69112, PROTEIN KINASE MOLECULES AND USES THEREFOR

Background of the Invention

[1258] Phosphate tightly associated with a molecule, e.g., a protein, has been known since the late nineteenth century. Since then, a variety of covalent linkages of phosphate to proteins have been found. The most common involve esterification of phosphate to serine, threonine, and tyrosine with smaller amounts being linked to lysine, arginine, histidine, aspartic acid, glutamic acid, and cysteine. The occurrence of phosphorylated molecules, e.g., proteins, implies the existence of one or more kinases, e.g., protein kinases, capable of phosphorylating various molecules, e.g., amino acid residues on proteins, and also of phosphatases, e.g., protein phosphatases, capable of hydrolyzing various phosphorylated molecules, e.g., phosphorylated amino acid residues on proteins.

[1259] Protein kinases play critical roles in the regulation of biochemical and morphological changes associated with cellular growth and division (D'Urso *et al.* (1990) *Science* 250:786-791; Birchmeier *et al.* (1993) *Bioessays* 15:185-189). For example, these kinases have been shown to participate in the transmission of signals from growth-factor receptors (Sturgill *et al.* (1988) *Nature* 344:715-718; Gomez *et al.* (1991) *Nature* 353:170-173), control of entry of cells into mitosis (Nurse (1990) *Nature* 344:503-508; Maller (1991) *Curr. Opin. Cell Biol.* 3:269-275), and regulation of actin bundling (Husain-Chishti *et al.* (1988) *Nature* 334:718-721). Protein kinases serve as growth factor receptors and signal transducers and have been implicated in cellular transformation and malignancy (Hunter *et al.* (1992) *Cell* 70:375-387; Posada *et al.* (1992) *Mol. Biol. Cell* 3:583-592; Hunter *et al.* (1994) *Cell* 79:573-582). Alterations in kinase genes and their products can lead to deregulated cell proliferation, a hallmark of cancer. Modulation of these genes and their regulatory activities may permit the control of tumor cell proliferation and invasion.

[1260] Protein kinases can be divided into different groups based on either amino acid sequence similarity or specificity for either serine/threonine or tyrosine residues. A small number of dual-specificity kinases have also been described. Within the broad classification, kinases can be further subdivided into families whose members share a higher degree of catalytic domain amino acid sequence identity and also have similar biochemical

properties. Most protein kinase family members also share structural features outside the kinase domain that reflect their particular cellular roles. These include regulatory domains that control kinase activity or interaction with other proteins (Hanks *et al.* (1988) *Science* 241:42-52).

[1261] Extracellular signal-regulated kinases/mitogen-activated protein kinases (ERKs\MAPKs) and cyclin-directed kinases (Cdks) represent two large families of serine-threonine kinases (see Songyang *et al.* (1996) *Mol. Cell. Biol.* 16: 6486-6493). Both types of kinases function in cell growth, cell division, and cell differentiation in response to extracellular stimuli. The ERK\MAPK family members are critical participants in intracellular signaling pathways. Upstream activators as well as the ERK\MAPK components are phosphorylated following contact of cells with growth factors or hormones or in response to cellular stressors, for example, heat, ultraviolet light, and inflammatory cytokines. These kinases transport messages that have been relayed from the plasma membrane to the cytoplasm by upstream kinases into the nucleus where they phosphorylate transcription factors and effect gene transcription modulation (Karin *et al.* (1995) *Curr. Biol.* 5: 747-757). Substrates of the ERK\MAPK family include c-fos, c-jun, APF2, and ETS family members Elk1, Sap1a, and c-Ets-1 (cited in Brott *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 963-968).

[1262] Signal-transduction pathways that employ members of the ERK/MAPK family of serine/threonine kinases are widely conserved among eukaryotes. The multiplicity of these pathways allows the cell to respond to divergent extracellular stimuli by initiating a broad array of responses ranging from cell growth to apoptosis. ERK/MAPK pathways are comprised of a three-tiered core-signaling module wherein ERK/MAPKs are regulated by MAPK/ERK kinases (MEKs), and MEKs, in turn, are regulated by MAPK kinase kinases (MAPKKKs). Mammalian stress-activated ERK/MAPK pathways have been implicated in numerous important physiological functions, including cell growth and proliferation, inflammatory responses, and apoptosis. For example, activation of the ERK1,2 signaling pathway by a mitogenic growth factor, a tumor promoter, or by transformation suppresses decorin gene expression in fibroblasts, which in turn may promote proliferation and migration of normal and malignant cells (Laine *et al.* (2000) *Biochem. J.* 349: 19-25).

[1263] Cdks regulate transitions between successive stages of the cell cycle. The activity of these molecules is controlled by phosphorylation events and by association with cyclin. Cdk activity is negatively regulated by the association of small inhibitory molecules

(Dymlacht (1997) *Nature* 389:148-152). Cdk targets include various transcriptional activators such as p110Rb, p107, and transcription factors, such as p53, E2F, and RNA polymerase II, as well as various cytoskeletal proteins and cytoplasmic signaling proteins (cited in Brott *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 963-968).

[1264] Protein kinases play critical roles in cellular growth, particularly in the transduction of signals for cell proliferation, differentiation, and apoptosis. Therefore, novel protein kinase polynucleotides and proteins are useful for modulating cellular growth, differentiation, and/or development.

Summary of the Invention

[1265] The present invention is based, in part, on the discovery of two novel human protein kinases, referred to herein as “68730” and “69112”. The nucleotide sequence of a cDNA encoding 68730 is shown in SEQ ID NO:32, and the amino acid sequence of a 68730 polypeptide is shown in SEQ ID NO:33. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:34. The nucleotide sequence of a cDNA encoding 69112 is shown in SEQ ID NO:35, and the amino acid sequence of a 69112 polypeptide is shown in SEQ ID NO:36. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:37.

[1266] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 68730 or 69112 protein or polypeptide, e.g., a biologically active portion of the 68730 or 69112 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:33 or 36. In other embodiments, the invention provides an isolated 68730 or 69112 nucleic acid molecule having the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37. In still other embodiments, the invention provides nucleic acid molecules that are sufficiently identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:32, 34, 35 or 37, wherein the nucleic acid encodes a full length 68730 or 69112 protein, or a biologically active fragment thereof.

[1267] In a related aspect, the invention further provides nucleic acid constructs which include a 68730 or 69112 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or

heterologous regulatory sequences. Also included are vectors and host cells containing 68730 or 69112 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing 68730 or 69112 nucleic acid molecules and polypeptides.

[1268] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 68730- or 69112-encoding nucleic acids.

[1269] In still another related aspect, isolated nucleic acid molecules that are antisense to a 68730 or 69112 encoding nucleic acid molecule are provided.

[1270] In another aspect, the invention features 68730 or 69112 polypeptides, and biologically active or antigenic fragments thereof, that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 68730- or 69112-mediated or related disorders. In another embodiment, the invention provides 68730 and 69112 polypeptides having a 68730 or 69112 activity. Preferred polypeptides are 68730 and 69112 proteins including at least one protein kinase family domain, and, preferably, having a 68730 or 69112 activity, e.g., a 68730 or 69112 activity as described herein.

[1271] In other embodiments, the invention provides 68730 and 69112 polypeptides, e.g., a 68730 polypeptide having the amino acid sequence shown in SEQ ID NO:33; a 69112 polypeptide having the amino acid sequence shown in SEQ ID NO:36; an amino acid sequence that is sufficiently identical to the amino acid sequence shown in SEQ ID NO:33 or 36; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:32, 34, 35 or 37, wherein the nucleic acid encodes a full length 68730 or 69112 protein, or a biologically active fragment thereof.

[1272] In a related aspect, the invention further provides nucleic acid constructs which include 68730 and 69112 nucleic acid molecules described herein.

[1273] In a related aspect, the invention provides 68730 and 69112 polypeptides or fragments operatively linked to non-68730 and non-69112 polypeptides to form fusion proteins.

[1274] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably, specifically bind 68730 or 69112 polypeptides.

[1275] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 68730 or 69112 polypeptides or nucleic acids.

[1276] In still another aspect, the invention provides a process for modulating 68730 or 69112 polypeptide or nucleic acid expression or activity, e.g. using the screened compounds. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 68730 or 69112 polypeptides or nucleic acids, such as conditions involving aberrant or deficient cellular proliferation or differentiation.

[1277] The invention also provides assays for determining the activity of or the presence or absence of 68730 or 69112 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[1278] In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 68730 or 69112 polypeptide or nucleic acid molecule in a biological sample (using, for example, 68730 or 69112 nucleic acid molecules or anti-68730 or anti-69112 antibodies), including for disease diagnosis.

[1279] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

[1280] **Detailed Description**

[1281] The human 68730 sequence (SEQ ID NO:32), which is approximately 1772 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1074 nucleotides (nucleotides 169 to 1242 of SEQ ID NO:32; nucleotides 1 to 1074 of SEQ ID NO:34), including the termination codon. The coding sequence encodes a 357 amino acid protein (SEQ ID NO:33).

[1282] The human 69112 sequence (SEQ ID NO:35), which is approximately 3579 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 1947 nucleotides (nucleotides 1213 to 3159 of SEQ ID NO:35; nucleotides 1 to 1947 of SEQ ID NO:37), including the termination codon. The coding sequence encodes a 648 amino acid protein (SEQ ID NO:36).

[1283] A hydropathy plot can be used to find hydrophobic and hydrophilic regions in 68730 and 69112 protein sequences (SEQ ID NO:33 and SEQ ID NO:36).

[1284] To determine whether a polypeptide or protein of interest has a conserved sequence or domain common to members of a protein family, the amino acid sequence of the protein can be searched against a database of HMMs (e.g., the Pfam database, release 2.1) using the default parameters. For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the Pfam database can be found in Sonhammer et al., (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov et al., (1990) *Meth. Enzymol.* 183:146-159; Gribskov et al., (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh et al., (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz et al., (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. See also, for example, *The HMMER User's Guide* available online. For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonhammer et al. (1997) *Protein* 28:405-420.

[1285] Using such search tools, the 68730 and 69112 protein sequences were found to contain significant structural characteristics in common with members of the protein kinase family of molecules. Some of these structural characteristics include, for example, a protein kinase domain, such as the consensus sequence of PFAM Accession No. PF00069 (which aligns with amino acids 23 to 279 of SEQ ID NO:33 (with a bit score of 346.3 and an E-value of 3.4e-100) and amino acids 356 to 613 of SEQ ID NO:36 (with a bit score of 350.2 and an E-value of 2.3e-101); and a serine/threonine kinase catalytic domain, such as the consensus sequence identified by SMART as Accession No. S_TKc (which aligns with amino acids 23 to 279 of SEQ ID NO:33 (with a bit score of 386.6 and an E-value of 2.5e-112) and amino acid residues 356 to 613 of SEQ ID NO:36 (with a bit score of 372.7 and a E-value of 3.8e-108). Another example of a common structural characteristic can include, for example, a tyrosine kinase domain, such as the consensus sequence of SMART Accession No. TyrKc (which aligns with amino acid residues 23 to 279 of SEQ ID NO:33 and amino acid residues 356 to 613 of SEQ ID NO:36). The bit score is 35.7 and the E-value is 4e-14 for SEQ ID NO:33. The bit score is 50.0 and the E-value is 5.3e-15 for SEQ ID NO:36.

[1286] As used herein, the term "protein kinase" includes a protein or polypeptide which is capable of modulating its own phosphorylation state or the phosphorylation state of

another molecule, e.g., protein or polypeptide. Protein kinases can have a specificity for (i.e., a specificity to phosphorylate) serine/threonine residues, tyrosine residues, or both serine/threonine and tyrosine residues, e.g., the dual specificity kinases. As used herein, a "protein kinase domain" includes a consensus sequence, e.g., PFAM Accession No. PF00069, that includes the catalytic domain of the enzyme. The catalytic domain can be characterized by the presence of an ATP binding signature sequence (e.g., Prosite Accession No. PS00107) and/or a serine/threonine kinase active-site signature sequence (e.g., Prosite Accession No. PS00108). The protein kinase domain of the present invention preferably includes a catalytic domain of about 150-400 amino acid residues in length, preferably about 200-300 amino acid residues in length, or more preferably about 225-300 amino acid residues in length, which includes at least one of the signature sequences or motifs described herein.

[1287] The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or signature sequence and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., rat or mouse proteins. Members of a family can also have common functional characteristics.

[1288] With regard to common structural characteristics described above, the protein kinases of the present invention include a protein kinase catalytic core or domain and can include at least one of the following signature sequences or motifs within the catalytic core: a protein kinase ATP-binding region signature sequence, a signature sequence that confers on the kinase its specificity for phosphorylation of either tyrosine or serine/threonine residues, and a tyrosine kinase phosphorylation site (See Hanks et al. (1988) *Science* 241:42-52).

[1289] The protein kinase ATP-binding region signature sequence is located in the N-terminal extremity of the catalytic domain and typically includes a glycine-rich stretch of residues in the vicinity of a lysine residue. A consensus sequence (Prosite Accession No. PS00107; SEQ ID NO:38) for this region is

[LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K.

[1290] In this and the following consensus sequence patterns, each element in the pattern is separated by a dash (-); square [] brackets indicate the particular residues that are accepted at that position; elaborate{ } brackets indicate the residues that are not accepted at that position; x indicates any residue is accepted at that position; repetition of a particular element is indicated by following the element with a numerical value or a numerical range enclosed in parentheses (i.e., above, x(5,18) indicates anywhere from 5 to 18 residues are present in the element, and any residue can be accepted at each of these 5 to 18 residue positions); and the standard IUPAC one-letter code for the amino acids is used. In the above consensus sequence pattern, lysine (K) binds ATP.

[1291] Analysis of the 68730 polypeptide for sequence patterns in the Prosite database showed a match to a protein kinase ATP binding signature pattern (Prosite Accession No. PS00107) at about amino acids 29 to 57 of SEQ ID NO:33 (LGTGAFSEVVLAEEKATGKLFAVKCIPKK; SEQ ID NO:42). The lysine residue at position 43, 47, 52, 56, and/or 57 can be involved in ATP binding. This protein kinase ATP-binding region signature sequence lies within a protein kinase domain spanning amino acid residues 23 to 279 of SEQ ID NO:33, as determined by a search using PFAM and SMART against the HMM database (HMMER 2.1.1).

[1292] Another region, located in the central part of the catalytic core or domain, contains a conserved aspartic acid residue, which is important for the catalytic activity of the enzyme (Knighton *et al.* (1991) *Science* 253:407-414). Two active-site signature sequences have been described for this region: one specific for serine/threonine kinases and one for tyrosine kinases. In both signature sequences aspartic acid (D) is conserved and is an active site residue. A consensus sequence for the serine/threonine kinases (Prosite Accession No. PS00108; SEQ ID NO:39) is [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3).

[1293] A consensus sequence for the tyrosine kinases (Prosite Accession No. PS00109; SEQ ID NO:40) is [LIVMFYC]-x-[HY]-x-D-[LIVMFY]-[RSTAC]-x(2)-N-[LIVMFYC](3).

[1294] Analysis of the 68730 polypeptide for sequence patterns in the Prosite database showed a match of IVHRDLKPENLLY (SEQ ID NO:43; amino acid residues 140 to 152 of SEQ ID NO:33) to the serine/threonine kinase active-site signature sequence shown above (PS00108; SEQ ID NO:39). The aspartic acid residue at position 144 of SEQ ID NO:33 is an active site residue. This serine/threonine kinase active-site signature

sequence lies within a protein kinase domain spanning amino acid residues 23 to 279 of SEQ ID NO:33, as determined by a search using PFAM and SMART against the HMM database (HMMER 2.1.1).

[1295] Analysis of the 69112 polypeptide for sequence patterns in the Prosite database showed a match of IVHRDLKPENLLV (SEQ ID NO:44; amino acid residues 473 to 485 of SEQ ID NO:36) to the serine/threonine kinase active-site signature sequence shown above (PS00108; SEQ ID NO:39). The aspartic acid residue at position 477 of SEQ ID NO:36 is an active site residue. This serine/threonine kinase active-site signature sequence lies within a protein kinase domain spanning amino acid residues 356 to 613 of SEQ ID NO:36, as determined by a search using PFAM and SMART against the HMM database (HMMER 2.1.1).

[1296] Substrates of tyrosine protein kinases are generally characterized by a lysine or an arginine seven residues to the N-terminal side of the phosphorylated tyrosine. An acidic residue (aspartic acid or glutamic acid) is often found at either three or four residues to the N-terminal side of the tyrosine (see Patschinsky *et al.* (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79:973-977; Hunter T. (1982) *J. Biol. Chem.* 257:4843-4848; Cooper *et al.* (1984) *J. Biol. Chem.* 259:7835-7841). Prosite Accession No. PS00007 is a consensus pattern for a tyrosine kinase phosphorylation site with the following sequence (SEQ ID NO:41): [RK]-x(2,3)-[DE]-x(2,3)-Y. In this sequence, tyrosine (Y) is the phosphorylation site.

[1297] Analysis of the 68730 polypeptide indicates that this sequence pattern occurs twice in the protein kinase domain spanning amino acid residues 23 to 279 of SEQ ID NO:33: in the sequence RIVEKGFY (SEQ ID NO:45) at amino acid residues 109 to 116, and in the sequence RQVLDAVYY (SEQ ID NO:46) at amino acid residues 126 to 134, wherein the phosphorylation site for these respective motifs reside at the tyrosine residue at positions 116 and 134.

[1298] The sequence analyses of the 68730 and 69112 proteins demonstrate that: 1) the 68730 protein can act as a serine/threonine protein kinase, and can also include at least one tyrosine kinase phosphorylation site, indicating that it can be phosphorylated by itself (i.e., autophosphorylation) or by another kinase; and 2) the 69112 protein is a serine/threonine protein kinase.

[1299] Isolated proteins of the present invention, preferably 68730 and 69112 proteins, have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:33 or 36, or are encoded by a nucleotide sequence sufficiently homologous or sufficiently

identical to SEQ ID NO:32, 34, 35 or 37. As used interchangeably herein, the terms "sufficiently homologous" and "sufficiently identical" refer to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently homologous.

[1300] As used interchangeably herein, a "68730 activity", "biological activity of 68730" or "functional activity of 68730", refers to an activity exerted by a 68730 protein, polypeptide or nucleic acid molecule on a 68730 responsive cell or a 68730 protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. As used interchangeably herein, a "69112 activity", "biological activity of 69112" or "functional activity of 69112", refers to an activity exerted by a 69112 protein, polypeptide or nucleic acid molecule on a 69112 responsive cell or a 69112 protein substrate, as determined *in vivo*, or *in vitro*, according to standard techniques.

[1301] Protein kinases play a role in signalling pathways associated with cellular growth. For example, protein kinases are involved in the regulation of signal transmission from cellular receptors, e.g., growth-factor receptors; entry of cells into mitosis; and the regulation of cytoskeleton function, e.g., actin bundling. Thus, the 68730 and 69112 molecules of the present invention can be involved in: 1) the regulation of transmission of signals from cellular receptors, e.g., cell growth factor receptors; 2) the modulation of the entry of cells, e.g., precursor cells, into mitosis; 3) the modulation of cellular differentiation; 4) the modulation of cell death; and 5) the regulation of cytoskeleton function, e.g., actin bundling. These kinases can function in these biological activities because of their ability to phosphorylate themselves or other substrate molecules.

[1302] Inhibition or over stimulation of the activity of protein kinases involved in signaling pathways associated with cellular growth can lead to perturbed cellular growth, which can in turn lead to cellular growth related disorders. As used herein, a "cellular growth related disorder" includes a disorder, disease, or condition characterized by a deregulation, e.g., an upregulation or a downregulation, of cellular growth. Cellular growth deregulation may be due to a deregulation of cellular proliferation, cell cycle progression, cellular differentiation and/or cellular hypertrophy. Examples of cellular growth related disorders include cardiovascular disorders such as heart failure, hypertension, atrial fibrillation, dilated cardiomyopathy, idiopathic cardiomyopathy, or angina; proliferative disorders or differentiative disorders such as cancer, e.g., melanoma, prostate cancer, cervical cancer, breast cancer, colon cancer, or sarcoma.

[1303] Accordingly, another embodiment of the invention features isolated 68730 and 69112 proteins and polypeptides having a 68730 or 69112 activity. Preferred proteins are 68730 and 69112 proteins having at least one protein kinase domain, at least a protein kinase ATP-binding signature sequence or a serine/threonine kinase active-site signature sequence, and, preferably, a 68730 or 69112 activity. Additional preferred 68730 and 69112 proteins have at least one protein kinase domain, at least one of a protein kinase ATP-binding signature sequence or a serine/threonine kinase active-site signature sequence, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:32, 34, 35 or 37, or a complement thereof.

[1304] The 68730 and 69112 polypeptides exhibit a domain shared by protein kinases and thus have similar biological activities. Accordingly, 68730 and 69112 can play a role (*e.g.*, involving a protein kinase domain) in cell proliferation and cancer, inflammation and apoptosis, and thus the 68730 and 69112 compositions of the invention (*e.g.*, nucleic acids, polypeptides, proteins, antibodies, and small molecule modulators of 68730 and 69112) can be used to modulate cell proliferation, *e.g.*, in cancer, inflammation or apoptosis, and furthermore can be used in screening assays to identify agents for modulating cell proliferation, as well as in detection or diagnostic assays to identify conditions involving aberrant cell proliferation.

[1305] As used herein, an "protein kinase-associated disorder" includes a disorder, disease or condition which is caused by, characterized by, or associated with a misregulation (*e.g.*, downregulation or upregulation) of a protein kinase-mediated activity or an abnormal

protein kinase-mediated activity. Protein kinase-associated disorders can detrimentally affect cellular functions such as cellular proliferation, growth, differentiation, or migration, cellular regulation of homeostasis, inter- or intra-cellular communication; tissue function, such as cardiac function or musculoskeletal function; systemic responses in an organism, such as nervous system responses, hormonal responses (*e.g.*, insulin response), or immune responses; and protection of cells from toxic compounds (*e.g.*, carcinogens, toxins, mutagens, and toxic byproducts of metabolic activity (*e.g.*, reactive oxygen species)). Accordingly, the 68730 and 69112 molecules of the invention, as protein kinases, can mediate various protein kinase-associated disorders, including cellular proliferative and/or differentiative disorders, hormonal disorders, immune and inflammatory disorders, neurological disorders, cardiovascular disorders, blood vessel disorders, and platelet disorders. As the 68730 and 69112 molecules of the invention can modulate protein kinase-mediated activities, they are useful for developing novel diagnostic and therapeutic agents for 68730- or 69112-mediated or related disorders, as described below.

[1306] Examples of cellular proliferative and/or differentiative disorders include cancer, *e.g.*, carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, *e.g.*, leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

[1307] As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, *i.e.*, an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, *i.e.*, characterizing or constituting a disease state, *e.g.*, malignant tumor growth, or may be categorized as non-pathologic, *i.e.*, a deviation from normal but not associated with a disease state, *e.g.*, cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory

system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[1308] The 68730 and 69112 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Typically, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L., (1991) *Crit. Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGL), Hodgkin's disease and Reed-Sternberg disease.

[1309] Protein kinase-associated disorders can include hormonal disorders, such as conditions or diseases in which the production and/or regulation of hormones in an organism is aberrant. Examples of such disorders and diseases include type I and type II diabetes mellitus, pituitary disorders (e.g., growth disorders), thyroid disorders (e.g., hypothyroidism or hyperthyroidism), and reproductive or fertility disorders (e.g., disorders which affect the organs of the reproductive system, e.g., the prostate gland, the uterus, or the vagina; disorders which involve an imbalance in the levels of a reproductive hormone in a subject;

disorders affecting the ability of a subject to reproduce; and disorders affecting secondary sex characteristic development, *e.g.*, adrenal hyperplasia).

[1310] Protein kinase-associated disorders also include immune disorders, such as autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency. Other examples of disorders include autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease (*e.g.*, Crohn's disease and ulcerative colitis), aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, respiratory inflammation (*e.g.*, asthma, allergic asthma, and chronic obstructive pulmonary disease), cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[1311] Additional protein kinase-associated disorders are neurological disorders. Such neurological disorders include, for example, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive

cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer's disease and Pick's disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington's disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma

multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

[1312] Cardiovascular disorders include, but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and

atresia, and aortic stenosis and atresia, disorders involving cardiac transplantation, and congestive heart failure.

[1313] Blood platelet disorders include, but are not limited to, thrombocytopenia due to a reduced number of megakaryocytes in the bone marrow, for example, as a result of chemotherapy; invasive disorders, such as leukemia, idiopathic or drug- or toxin-induced aplasia of the marrow, or rare hereditary amegakaryocytic thrombocytopenias; ineffective thrombopoiesis, for example, as a result of megaloblastic anemia, alcohol toxicity, vitamin B12 or folate deficiency, myelodysplastic disorders, or rare hereditary disorders (e.g., Wiskott-Aldrich syndrome and May-hegglin anomaly); a reduction in platelet distribution, for example, as a result of cirrhosis, a splenic invasive disease (e.g., Gaucher's disease), or myelofibrosis with extramedullary myeloid metaplasia; increased platelet destruction, for example, as a result of removal of IgG-coated platelets by the mononuclear phagocytic system (e.g., idiopathic thrombocytopenic purpura (ITP), secondary immune thrombocytopenia (e.g., systemic lupus erythematosus, lymphoma, or chronic lymphocytic leukemia), drug-related immune thrombocytopenias (e.g., as with quinidine, aspirin, and heparin), post-transfusion purpura, and neonatal thrombocytopenia as a result of maternal platelet autoantibodies or maternal platelet alloantibodies). Also included are thrombocytopenia secondary to intravascular clotting and thrombin induced damage to platelets as a result of, for example, obstetric complications, metastatic tumors, severe gram-negative bacteremia, thrombotic thrombocytopenic purpura, or severe illness. Also included is dilutional thrombocytopenia, for example, due to massive hemorrhage. Blood platelet disorders also include, but are not limited to, essential thrombocytosis and thrombocytosis associated with, for example, splenectomy, acute or chronic inflammatory diseases, hemolytic anemia, carcinoma, Hodgkin's disease, lymphoproliferative disorders, and malignant lymphomas.

[1314] Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger

disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi's sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

[1315] The 68730 and 69112 proteins, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:33 and 36, are collectively referred to as "polypeptides or proteins of the invention" or "68730 and 69112 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "68730 and 69112 nucleic acids." 68730 molecules refer to 68730 nucleic acids, polypeptides, and antibodies, and 69112 molecules refer to 69112 nucleic acids, polypeptides, and antibodies.

[1316] As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[1317] The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of

the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be sufficiently free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[1318] As used herein, the term "hybridizes under stringent conditions" describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Aqueous and nonaqueous methods are described in that reference and either can be used. A preferred example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Another preferred stringent hybridization condition is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5M Sodium Phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:32, or SEQ ID NO:35, corresponds to a naturally-occurring nucleic acid molecule.

[1319] As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[1320] As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 68730 or 69112 protein, preferably a mammalian 68730 or 69112 protein, and can further include non-coding regulatory sequences, and introns.

[1321] An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which

the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means a preparation of 68730 or 69112 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-68730 or non-69112 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-68730 or non-69112 chemicals. When the 68730 or 69112 protein, or biologically active portion thereof, is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[1322] A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 68730 or 69112 (e.g., the sequence of SEQ ID NO:32, 34, 35 or 37) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the protein kinase domain, are predicted to be particularly unamenable to alteration.

[1323] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 68730 or 69112 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 68730 or 69112 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 68730 or 69112 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID

NO:32, 34, 35 or 37, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[1324] Biologically active portions of a 68730 or 69112 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 68730 or 69112 protein, e.g., the amino acid sequence shown in SEQ ID NO:33 or 36, which include less amino acids than the full length 68730 or 69112 proteins, and exhibit at least one activity of a 68730 or 69112 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 68730 or 69112 protein, e.g., protein kinase activity. A biologically active portion of a 68730 or 69112 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 68730 or 69112 protein can be used as targets for developing agents which modulate a 68730 or 69112 mediated activity, e.g., protein kinase activity.

[1325] Calculations of homology or sequence identity between sequences (the terms are used interchangeably herein) are performed as follows.

[1326] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 68730 amino acid sequence of SEQ ID NO:33 having 357 amino acid residues, at least 107, preferably at least 143, more preferably at least 179, even more preferably at least 214, and even more preferably at least 250, 286, 321, or 357 amino acid residues are aligned, or when aligning a second sequence to the 69112 amino acid sequence of SEQ ID NO:36 having 648 amino acid residues, at least 194, preferably at least 259, more preferably at least 324, even more preferably at least 389, and even more preferably at least 454, 518, 583, or 648 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that

position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[1327] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) is using a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[1328] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[1329] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al., (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 68730 or 69112 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 68730 or 69112 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic*

Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[1330] "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[1331] "Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

[1332] A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[1333] Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

[1334] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 68730 or 69112 polypeptide described herein, e.g., a full length 68730 or 69112 protein, or a fragment thereof, e.g., a biologically active portion of 68730 or 69112 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify nucleic acid molecule encoding a polypeptide of the invention, 68730 and 69112 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

[1335] In one embodiment, an isolated nucleic acid molecule of the present invention includes the nucleotide sequence shown in SEQ ID NO:32, or a portion of this nucleotide sequence. In one embodiment, the nucleic acid molecule includes sequences encoding the human 68730 protein (*i.e.*, "the coding region", from nucleotides 169 to 1242 of SEQ ID NO:32, including the termination codon), as well as untranslated (*e.g.*, noncoding) sequences, *e.g.*, 5' untranslated sequence (*i.e.*, nucleotides 1 to 168 of SEQ ID NO:32) and/or 3' untranslated sequence (nucleotides 1243 to 1772 of SEQ ID NO:32). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:32 (*e.g.*, nucleotides 169 to 1242 of SEQ ID NO:32, corresponding to nucleotides 1 to 1074 SEQ ID NO:34) and, *e.g.*, no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the mature protein of SEQ ID NO:33.

[1336] In another embodiment, an isolated nucleic acid molecule of the present invention includes the nucleotide sequence shown in SEQ ID NO:35, or a portion of this nucleotide sequence. In one embodiment, the nucleic acid molecule includes sequences encoding the human 69112 protein (*i.e.*, "the coding region", from nucleotides 1213 to 3159 of SEQ ID NO:35, including the termination codon), as well as untranslated (*e.g.*, noncoding) sequences, *e.g.*, 5' untranslated sequence (*i.e.*, nucleotides 1 to 1212 of SEQ ID NO:35) and/or 3' untranslated sequence (*i.e.*, nucleotides 3160 to 3579 of SEQ ID NO:35). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:35 (*e.g.*, nucleotides 1213 to 3159 of SEQ ID NO:35, corresponding to nucleotides 1 to 1947 SEQ ID NO:37) and, *e.g.*, no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to the mature protein of SEQ ID NO:36.

[1337] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently homologous to the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37, respectively, thereby forming a stable duplex.

[1338] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 60%, 65%, 70%, 75%, 80%,

85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37. In the case of an isolated nucleic acid molecule which is longer than or equivalent in length to the reference sequence, *e.g.*, SEQ ID NO:32, 34, 35 or 37, the comparison is made with the full length of the reference sequence. Where the isolated nucleic acid molecule is shorter than the reference sequence, *e.g.*, shorter than SEQ ID NO:32, 34, 35 or 37, the comparison is made to a segment of the reference sequence of the same length (excluding any loop required by the homology calculation).

68730 and 69112 Nucleic Acid Fragments

[1339] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:32, 34, 35 or 37. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 68730 or 69112 protein, *e.g.*, an immunogenic or biologically active portion of a 68730 or 69112 protein. A fragment can comprise: nucleotides 234 to 1005 of SEQ ID NO:32, which encodes a protein kinase domain of human 68730, *e.g.*, amino acid residues 23 to 279 of SEQ ID NO:33; or nucleotides 2278 to 3051 of SEQ ID NO:35, which encodes a protein kinase domain of human 69112, *e.g.*, amino acid residues 356 to 613 of SEQ ID NO:36. The nucleotide sequences determined from the cloning of the 68730 and 69112 genes allow for the generation of probes and primers designed for use in identifying and/or cloning other family members, or fragments thereof, as well as homologues, or fragments thereof, of 68730 and 69112 genes from other species.

[1340] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' untranslated region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 150 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not be construed as encompassing those fragments that may have been disclosed prior to the invention.

[1341] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or

more domain, region, or functional site described herein. Thus, for example, the nucleic acid fragment can include a protein kinase domain. In a preferred embodiment the fragment is at least, 50, 100, 200, 300, 400, 500, 600, 700, or 900 base pairs in length.

[1342] 68730 and 69112 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:32, 34, 35 or 37, or of a naturally occurring allelic variant or mutant of SEQ ID NO:32, 34, 35 or 37.

[1343] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[1344] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes a protein kinase domain (e.g., at about nucleotides 234 to 1005 of SEQ ID NO:32, and at about nucleotides 2278 to 3051 of SEQ ID NO:35), or a fragment thereof.

[1345] In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of a 68730 or 69112 sequence, *e.g.*, a region described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: a protein kinase ATP binding domain (*e.g.*, at about nucleotides 253 to 339 of SEQ ID NO:32); and a serine/threonine protein kinase active site signature sequence (*e.g.*, at about nucleotides 586 to 624 of SEQ ID NO:32 and at about nucleotides 2629 to 2667 of SEQ ID NO:35).

[1346] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[1347] A nucleic acid fragment encoding a "biologically active portion of a 68730 polypeptide" or a "biologically active portion of a 69112 polypeptide" can be prepared by

isolating a portion of the nucleotide sequence of SEQ ID NO:32, 34, 35 or 37, which encodes a polypeptide having a 68730 or 69112 biological activity (e.g., the biological activities of the 68730 or 69112 proteins as described herein), expressing the encoded portion of the 68730 or 69112 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 68730 or 69112 protein. For example, a nucleic acid fragment encoding a biologically active portion of 68730 or 69112 includes a protein kinase domain (e.g., at about nucleotides 235 to 1005 of SEQ ID NO:32 or nucleotides 2278 to 3051 of SEQ ID NO:35). A nucleic acid fragment encoding a biologically active portion of a 68730 or 69112 polypeptide, can comprise a nucleotide sequence which is greater than 200-1200 or more nucleotides in length.

[1348] In preferred embodiments, nucleic acids include a nucleotide sequence which is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:32, 34, 35 or 37, or a complement thereof.

68730 and 69112 Nucleic Acid Variants

[1349] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:32, 34, 35 or 37. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 68730 or 69112 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:33 or 36. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[1350] Nucleic acids of the present invention can be chosen for having codons, which are preferred, or non preferred, for a particular expression system. For example, the nucleic acid can be one in which at least one codon, and preferably at least 10%, or 20% of the codons have been altered such that the sequence is optimized for expression in bacterial (e.g., *E. coli*), yeast, human, insect, or nonmammalian cells (e.g., Chinese hamster ovary cells).

[1351] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared to the encoded product).

[1352] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:32, 34, 35 or 37, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[1353] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the amino acid sequence shown in SEQ ID NO:33 or a fragment of this sequence. Such nucleic acid molecules can readily be obtained as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:34 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 68730 and 69112 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 68730 or 69112 gene. Preferred variants include those that are correlated with protein kinase activity.

[1354] Allelic variants of 68730 and 69112, e.g., human 68730 and 69112, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 68730 and 69112 proteins within a population that maintain the ability to modulate the phosphorylation state of itself or another protein or polypeptide. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:33 or 36, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 68730 and 69112, e.g., human 68730 and human 69112, protein within a population that do not have the ability to activate signal transduction. Non-functional allelic variants will typically contain a non-conservative

substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:33 or 36, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

[1355] Moreover, nucleic acid molecules encoding other 68730 and 69112 family members and, thus, which have a nucleotide sequence which differs from the 68730 and 69112 sequences of SEQ ID NO:32, 34, 35 or 37 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 68730 and 69112 Nucleic Acid Molecules

[1356] In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 68730 or 69112. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 68730 or 69112 coding strand, or to only a portion thereof (e.g., the coding region of human 68730 corresponding to SEQ ID NO:34 or the coding region of human 69112 corresponding to SEQ ID NO:37). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 68730 or 69112 (e.g., the 5' and 3' untranslated regions).

[1357] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 68730 or 69112 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 68730 or 69112 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 68730 or 69112 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[1358] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex

formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[1359] The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 68730 or 69112 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[1360] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al., (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-O-methylribonucleotide (Inoue et al., (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al., (1987) *FEBS Lett.* 215:327-330).

[1361] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 68730- or 69112-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 68730 or 69112 cDNA disclosed herein (i.e., SEQ ID NO:32, 34, 35 or 37), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the

active site is complementary to the nucleotide sequence to be cleaved in a 68730- or 69112-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, 68730 or 69112 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[1362] 68730 and 69112 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 68730 or 69112 (e.g., the 68730 or 69112 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 68730 or 69112 gene in target cells. See generally, Helene, C., (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al., (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J., (1992) *Bioassays* 14(12):807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[1363] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[1364] A 68730 and a 69112 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. et al., (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. et al., (1996) *supra*; Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675.

[1365] PNAs of 68730 and 69112 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription

or translation arrest or inhibiting replication. PNAs of 68730 and 69112 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B., (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. et al., (1996) *supra*; Perry-O'Keefe *supra*).

[1366] In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al., (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon, (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[1367] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 68730 or 69112 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 68730 or 69112 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

Isolated 68730 and 69112 Polypeptides

[1368] In another aspect, the invention features, isolated 68730 and 69112 proteins, or fragments, e.g., biologically active portions, for use as immunogens or antigens to raise or test (or more generally to bind) anti-68730 and anti-69112 antibodies. 68730 and 69112 proteins can be isolated from cells or tissue sources using standard protein purification techniques. 68730 and 69112 proteins, or fragments thereof, can be produced by recombinant DNA techniques or synthesized chemically.

[1369] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing

events, and alternative translational and postranslational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in a polypeptide with substantially the same postranslational modifications as the polypeptide expressed in a native host cell, or in systems which result in the alteration or omission of postranslational modifications, e.g., glycosylation or cleavage, which otherwise occur when the polypeptide is expressed in its native host cell.

[1370] In a preferred embodiment, a 68730 or 69112 polypeptide has one or more of the following characteristics: i) the ability to act as a protein kinase or to activate a protein kinase activity; ii) the ability to act as a substrate for a protein kinase, *e.g.*, a tyrosine kinase; iii) a molecular weight, *e.g.*, a deduced molecular weight, amino acid composition or other physical characteristic of the polypeptide of SEQ ID NO:33 or 36; iv) an overall sequence similarity of at least 50%, preferably at least 60%, more preferably at least 70, 80, 90, or 95%, with a polypeptide of SEQ ID NO:33 or 36; v) a protein kinase domain which preferably has an overall sequence similarity of about 70%, 80%, 90% or 95% with amino acid residues 23 to 279 of SEQ ID NO:33 or amino acid residues 356 to 613 of SEQ ID NO:36; and vi) the presence of at least 70%, preferably 80%, and most preferably 95% of the cysteines found in the amino acid sequence of the 68730 or 69112 native protein.

[1371] In a preferred embodiment, the 68730 or 69112 protein, or a fragment thereof, differs from the corresponding sequence in SEQ ID NO:33 or 36. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:33 or 36 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:33 or 36. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the protein kinase domain. In another preferred embodiment one or more differences are in non-active site residues, *e.g.* outside of the protein kinase domain.

[1372] Other embodiments include a protein that contain one or more changes in amino acid sequence, *e.g.*, a change in an amino acid residue which is not essential for activity. Such 68730 and 69112 proteins differ in amino acid sequence from SEQ ID NO:33 and 36, respectively, yet retain biological activity.

[1373] In one embodiment, a biologically active portion of a 68730 and 69112 protein includes a protein kinase domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of a native 68730 and 69112 protein.

[1374] In a preferred embodiment, the 68730 protein has an amino acid sequence shown in SEQ ID NO:33. In other embodiments, the 68730 protein is sufficiently identical to SEQ ID NO:33. In yet another embodiment, the 68730 protein is sufficiently identical to SEQ ID NO:33 and retains the biological activity of the protein of SEQ ID NO:33, as described in detail above. Accordingly, in another embodiment, the 68730 protein is a protein which includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:33.

[1375] In another preferred embodiment, the 69112 protein has an amino acid sequence shown in SEQ ID NO:36. In other embodiments, the 69112 protein is sufficiently identical to SEQ ID NO:36. In yet another embodiment, the 69112 protein is sufficiently identical to SEQ ID NO:36 and retains the biological activity of the protein of SEQ ID NO:36, as described in detail above. Accordingly, in another embodiment, the 68730 protein is a protein which includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:36.

68730 and 69112 Chimeric or Fusion Proteins

[1376] In another aspect, the invention provides 68730 and 69112 chimeric or fusion proteins. As used herein, a 68730 and 69112 "chimeric protein" or "fusion protein" includes a 68730 or 69112 polypeptide linked to a non-68730 or non-69112 polypeptide. A "non-68730 polypeptide" and "non-69112 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not sufficiently homologous to the 68730 or 69112 protein, e.g., a protein which is different from the 68730 or 69112 protein and which is derived from the same or a different organism. The 68730 or polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 68730 or 69112 amino acid sequence. In a preferred embodiment, a 68730 or 69112 fusion protein includes at least one (or two) biologically active portion of a 68730 or 69112 protein. The non-68730 or non-69112 polypeptide can be fused to the N-terminus or C-terminus of the 68730 or 69112 polypeptide.

[1377] The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-68730 or GST-69112 fusion protein in which the 68730 or 69112 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 68730 or 69112. Alternatively, the fusion protein can be a 68730 or 69112 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of 68730 or 69112 can be increased through use of a heterologous signal sequence.

[1378] Fusion proteins can include all or a part of a serum protein, *e.g.*, a portion of an immunoglobulin (*e.g.*, IgG, IgA, or IgE), *e.g.*, an Fc region and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

[1379] The 68730 and 69112 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 68730 and 69112 fusion proteins can be used to affect the bioavailability of a 68730 or 69112 substrate. 68730 and 69112 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 68730 or 69112 protein; (ii) mis-regulation of the 68730 or 69112 gene; and (iii) aberrant post-translational modification of a 68730 or 69112 protein.

[1380] Moreover, the 68730 and 69112 fusion proteins of the invention can be used as immunogens to produce anti-68730 and anti-69112 antibodies in a subject, to purify 68730 and 69112 ligands, and in screening assays, to identify molecules which inhibit the interaction of 68730 with a 68730 substrate and 69112 with a 69112 substrate.

[1381] Expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A 68730- or 69112-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 68730 or 69112 protein.

Variants of 68730 and 69112 Proteins

[1382] In another aspect, the invention also features a variant of a 68730 and a 69112 polypeptide, *e.g.*, which functions as an agonist (mimetics) or as an antagonist. Variants of the 68730 and 69112 proteins can be generated by mutagenesis, *e.g.*, discrete point mutation, the insertion or deletion of sequences or the truncation of a 68730 or a 69112 protein. An agonist of the 68730 and 69112 proteins can retain sufficiently the same, or a

subset, of the biological activities of the naturally occurring form of a 68730 or 69112 protein. An antagonist of a 68730 or 69112 protein can inhibit one or more of the activities of the naturally occurring form of the 68730 or 69112 protein by, for example, competitively modulating a 68730- or 69112-mediated activity of a 68730 or 69112 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 68730 or 69112 protein.

[1383] Variants of a 68730 or 69112 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 68730 or 69112 protein for agonist or antagonist activity.

[1384] Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 68730 or 69112 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 68730 or 69112 protein.

[1385] Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[1386] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Such methods are adaptable for rapid screening of the gene libraries generated by combinatorial mutagenesis of 68730 and 69112 proteins. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 68730 and 69112 variants (Arkin and Yourvan, (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al., (1993) *Protein Engineering* 6(3):327-331).

[1387] Cell based assays can be exploited to analyze a variegated 68730 or 69112 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 68730 or 69112 in a substrate-dependent manner. The transfected cells are then contacted with 68730 or 69112 and the effect of the expression of the mutant on signaling by the respective 68730 or 69112 substrate can be detected, e.g., by measuring protein kinase activity. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 68730 or 69112 substrate, and the individual clones further characterized.

[1388] In another aspect, the invention features a method of making a 68730 and a 69112 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 68730 or 69112 polypeptide, e.g., a naturally occurring 68730 or 69112 polypeptide. The method includes: altering the sequence of a 68730 or 69112 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[1389] In another aspect, the invention features a method of making a fragment or analog of a 68730 and a 69112 polypeptide having a biological activity of a naturally occurring 68730 or 69112 polypeptide. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of a 68730 or 69112 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-68730 and Anti-69112 Antibodies

[1390] In another aspect, the invention provides an anti-68730 and an anti-69112 antibody. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, Fab and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin.

[1391] The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric, humanized, fully human, non-human (e.g., murine, rat, rabbit, or goat), or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

[1392] A full-length 68730 or 69112 protein, or an antigenic peptide fragment of 68730 or 69112, can be used as an immunogen or can be used to identify anti-68730 or anti-69112 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptides of 68730 and 69112 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:33 or 36 and encompass an epitope of 68730 or 69112, respectively. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more

preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[1393] Fragments of 68730 and 69112 which include, e.g., residues 23 to 279 of SEQ ID NO:33 and residues 356 to 613 of SEQ ID NO:36, respectively, can be used as immunogens to make an antibody against what is believed to be the protein kinase region of the 68730 and 69112 protein.

[1394] Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

[1395] In an alternative embodiment the antibody fails to bind to an Fc receptor, e.g., it is a type which does not support Fc receptor binding or has been modified, e.g., by deletion or other mutation, such that it does not have a functional Fc receptor binding region.

[1396] Preferred epitopes encompassed by the antigenic peptide are regions of 68730 and 69112 which are located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 68730 and 69112 protein sequences can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 68730 and 69112 protein, respectively, and are thus likely to constitute surface residues useful for targeting antibody production.

[1397] In a preferred embodiment the antibody binds an epitope on any domain or region on 68730 and 69112 proteins described herein.

[1398] Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients.

[1399] The anti-68730 and anti-69112 antibodies can each be a single chain antibody. A single-chain antibody (scFV) may be engineered as described, for example, in Colcher, D. et al., *Ann. NY Acad. Sci.* 1999 Jun 30;880:263-80; and Reiter, Y., *Clin. Cancer Res.* 1996 Feb;2(2):245-52. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 68730 and 69112 protein.

[1400] Anti-68730 and anti-69112 antibodies (e.g., monoclonal antibodies) can be used to isolate 68730 and 69112, respectively, by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-68730 or anti-69112 antibody

can be used to detect 68730 or 69112 protein, respectively, (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-68730 and anti-69112 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[1401] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

[1402] A vector can include a 68730 or 69112 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be

introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 68730 and 69112 proteins, mutant forms of 68730 and 69112 proteins, fusion proteins, and the like).

[1403] The recombinant expression vectors of the invention can be designed for expression of 68730 or 69112 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in bacterial cells (e.g., *E. coli*), insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[1404] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S., (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[1405] Purified fusion proteins can be used in 68730 and 69112 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 68730 and 69112 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[1406] To maximize recombinant protein expression in *E. coli* is to express the protein in host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[1407] The 68730 and 69112 expression vectors can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector, or a vector suitable for expression in mammalian cells.

[1408] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[1409] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton, (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore, (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al., (1983) *Cell* 33:729-740; Queen and Baltimore, (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al., (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss, (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman, (1989) *Genes Dev.* 3:537-546).

[1410] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue

specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

[1411] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 68730 or 69112 nucleic acid molecule within a recombinant expression vector or a 68730 or 69112 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[1412] A host cell can be any prokaryotic or eukaryotic cell. For example, a 68730 or 69112 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[1413] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[1414] A host cell of the invention can be used to produce (i.e., express) a 68730 or a 69112 protein. Accordingly, the invention further provides methods for producing 68730 and 69112 proteins using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 68730 or 69112 protein has been introduced) in a suitable medium such that a 68730 or a 69112 protein is produced. In another embodiment, the method further includes isolating a 68730 or a 69112 protein from the medium or the host cell.

[1415] In another aspect, the invention features, a cell or purified preparation of cells which includes a 68730 and a 69112 transgene, or which otherwise misexpresses 68730 or 69112. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 68730 and a 69112 transgene, e.g., a heterologous form of a 68730 and a 69112, e.g., a gene derived from humans (in the case of a non-human cell). The 68730 or 69112 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous 68730 or 69112, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed 68730 or 69112 alleles or for use in drug screening.

[1416] In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 68730 or 69112 polypeptide.

[1417] Also provided are cells or a purified preparation thereof, e.g., human cells, in which an endogenous 68730 or 69112 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 68730 or 69112 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 68730 or 69112 gene. For example, an endogenous 68730 or 69112 gene, e.g., a gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published on May 16, 1991.

Transgenic Animals

[1418] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 68730 or 69112 protein and for identifying and/or evaluating modulators of 68730 or 69112 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 68730 or 69112 gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

[1419] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 68730 or 69112 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 68730 or 69112 transgene in its genome and/or expression of 68730 or 69112 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 68730 or 69112 protein can further be bred to other transgenic animals carrying other transgenes.

[1420] 68730 and 69112 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[1421] The invention also includes a population of cells from a transgenic animal, as discussed herein.

Uses

[1422] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

[1423] The isolated nucleic acid molecules of the invention can be used, for example, to express a 68730 or 69112 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 68730 or 69112 mRNA (e.g., in a biological sample) or a genetic alteration in a 68730 or a 69112 gene, and to modulate 68730 or 69112 activity, as described further below. The 68730 and 69112 proteins can be used to treat disorders characterized by insufficient or excessive production of a 68730 or 69112 substrate or production of 68730 or 69112 inhibitors. In addition, the 68730 and 69112 proteins can be used to screen for naturally occurring 68730 and 69112 substrates, to screen for drugs or compounds which modulate 68730 or 69112 activity, as well as to treat disorders characterized by insufficient or excessive production of 68730 or 69112 protein or production of 68730 or 69112 protein forms which have decreased, aberrant or unwanted activity compared to 68730 or 69112 wild-type protein. Such disorders include those characterized by aberrant signaling or aberrant, e.g., hyperproliferative, cell growth. Moreover, the anti-68730 and anti-69112 antibodies of the invention can be used to detect and isolate 68730 or 69112 proteins, regulate the bioavailability of 68730 or 69112 proteins, and modulate 68730 or 69112 activity.

[1424] A method of evaluating a compound for the ability to interact with, e.g., bind, subject 68730 and 69112 polypeptides are provided. The method includes: contacting the compound with the subject 68730 or 69112 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 68730 or 69112 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 68730 or 69112 polypeptides. It can also be used to find natural or synthetic inhibitors of subject 68730 or 69112 polypeptides. Screening methods are discussed in more detail below.

Screening Assays

[1425] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 68730 or 69112 proteins, have a stimulatory or inhibitory effect on, for example, 68730 or 69112 expression, or 68730 or 69112 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 68730 or 69112 substrate. Compounds thus

identified can be used to modulate the activity of target gene products (e.g., 68730 or 69112 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[1426] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 68730 or 69112 protein or polypeptide, or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 68730 or 69112 protein or polypeptide or a biologically active portion thereof.

[1427] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries [libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive] (see, e.g., Zuckermann, R.N. et al., *J. Med. Chem.* 1994, 37: 2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[1428] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.*, (1994). *J. Med. Chem.* 37:2678; Cho *et al.*, (1993) *Science* 261:1303; Carrell *et al.*, (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.*, (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.*, (1994) *J. Med. Chem.* 37:1233.

[1429] Libraries of compounds may be presented in solution (e.g., Houghten, (1992) *Biotechniques* 13:412-421), or on beads (Lam, (1991) *Nature* 354:82-84), chips (Fodor, (1993) *Nature* 364:555-556), bacteria or spores (Ladner, United States Patent No. 5,223,409), plasmids (Cull *et al.*, (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (Scott and Smith, (1990) *Science* 249:386-390); (Devlin, (1990) *Science* 249:404-406); (Cwirla *et al.*, (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici, (1991) *J. Mol. Biol.* 222:301-310); (Ladner, *supra*).

[1430] In another embodiment, an assay is a cell-based comprising contacting a cell expressing a 68730 or 69112 target molecule (*e.g.*, a 68730 or 69112 phosphorylation substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the 68730 or 69112 target molecule. Determining the ability of the test compound to modulate the activity of a 68730 or 69112 target molecule can be accomplished, for example, by determining the ability of the 68730 or 69112 protein to bind to or interact with the 68730 or 69112 target molecule, or by determining the ability of the 68730 or 69112 protein to phosphorylate the 68730 or 69112 target molecule.

[1431] The ability of the 68730 or 69112 protein to phosphorylate a substrate/target molecule can be determined by, for example, an *in vitro* kinase assay. Briefly, a substrate/target molecule, *e.g.*, an immunoprecipitated substrate/target molecule from a cell line expressing such a molecule, can be incubated with the 68730 or 69112 protein and radioactive ATP, *e.g.*, [γ -³²P] ATP, in a buffer containing MgCl₂ and MnCl₂, *e.g.*, 10 mM MgCl₂ and 5 mM MnCl₂. Following the incubation, the immunoprecipitated substrate/target molecule can be separated by SDS-polyacrylamide gel electrophoresis under reducing conditions, transferred to a membrane, *e.g.*, a PVDF membrane, and autoradiographed. The appearance of detectable bands on the autoradiograph indicates that the 68730 or 69112 substrate/target molecule has been phosphorylated. Phosphoaminoacid analysis of the phosphorylated substrate/target molecule can also be performed in order to determine which residues on the 68730 or 69112 substrate/target molecule are phosphorylated. Briefly, the radiophosphorylated protein band can be excised from the SDS gel and subjected to partial acid hydrolysis. The products can then be separated by one-dimensional electrophoresis and analyzed on, for example, a phosphoimager and compared to ninhydrin-stained phosphoaminoacid standards.

[1432] The ability of the test compound to modulate 68730 or 69112 binding to a compound, *e.g.*, a 68730 or 69112 substrate, or to bind to 68730 or 69112 can also be evaluated. This can be accomplished, for example, by coupling the compound, *e.g.*, the substrate, with a radioisotope or enzymatic label such that binding of the compound, *e.g.*, the substrate, to 68730 or 69112 can be determined by detecting the labeled compound, *e.g.*, substrate, in a complex. Alternatively, 68730 or 69112 can be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 68730 or 69112 binding to a 68730 or 69112 substrate in a complex. For example, compounds (*e.g.*,

substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[1433] The ability of a compound (e.g., a 68730 or 69112 substrate) to interact with 68730 or 69112 with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 68730 or 69112 without the labeling of either the compound or 68730 or 69112.

McConnell, H. M. et al., (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 68730 or 69112.

[1434] In a preferred embodiment, determining the ability of the 68730 or 69112 protein to bind to or interact with a 68730 or 69112 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+} , diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., chloramphenicol acetyl transferase), or detecting a target-regulated cellular response.

[1435] In yet another embodiment, a cell-free assay is provided in which a 68730 or 69112 protein, or biologically active portion thereof, is contacted with a test compound and the ability of the test compound to bind to the 68730 or 69112 protein, or biologically active portion thereof is evaluated. Preferred biologically active portions of the 68730 and 69112 proteins to be used in assays of the present invention include fragments which participate in interactions with non-68730 or non-69112 molecules, e.g., fragments with high surface probability scores.

[1436] Soluble and/or membrane-bound forms of isolated proteins (e.g., 68730 and 69112 proteins, or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to

utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[1437] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[1438] In one embodiment, assays are performed where the ability of an agent to block protein kinase activity within a cell is evaluated.

[1439] The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, et al., U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[1440] In another embodiment, determining the ability of the 68730 or 69112 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C., (1991) *Anal. Chem.* **63**:2338-2345 and Szabo et al., (1995) *Curr. Opin. Struct. Biol.* **5**:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the

optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[1441] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[1442] It may be desirable to immobilize either 68730 or 69112, an anti-68730 or anti-69112 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 68730 or 69112 protein, or interaction of a 68730 or 69112 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/68730 or glutathione-S-transferase/69112 fusion proteins or glutathione-S-transferase/68730 target or glutathione-S-transferase/69112 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed 68730 or 69112 target protein or 68730 or 69112 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 68730 or 69112 binding or activity determined using standard techniques.

[1443] Other techniques for immobilizing either a 68730 or 69112 protein or a 68730 or 69112 target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 68730 or 69112 protein or 68730 or 69112 target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96-well plates (Pierce Chemical).

[1444] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[1445] In one embodiment, this assay is performed utilizing antibodies reactive with 68730 or 69112 protein or 68730 or 69112 target molecules but which do not interfere with binding of the 68730 or 69112 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 68730 or 69112 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 68730 or 69112 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 68730 or 69112 protein or 68730 or 69112 target molecule.

[1446] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., *Trends Biochem Sci* 1993 Aug; 18(8):284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. et al., eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. et al., eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., *J Mol. Recognit.* 1998 Winter; 11(1-6):141-8; Hage, D.S., and Tweed, S.A., *J. Chromatogr. B Biomed. Sci. Appl.* 1997 Oct 10; 699 (1-2):499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[1447] In a preferred embodiment, the assay includes contacting the 68730 or 69112 protein, or a biologically active portion thereof, with a known compound which binds 68730 or 69112 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 68730 or 69112 protein, wherein determining the ability of the test compound to interact with a 68730 or 69112 protein includes determining the ability of the test compound to preferentially bind to 68730 or 69112, or a biologically active portion thereof, or to modulate the activity of a 68730 or 69112 target molecule, as compared to the known compound.

[1448] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 68730 and 69112 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 68730 or 69112 protein through modulation of the activity of a downstream effector of a 68730 or a 69112 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[1449] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within

reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[1450] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[1451] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[1452] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of

reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[1453] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[1454] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[1455] In yet another aspect, the 68730 and 69112 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al., (1993) *Cell* 72:223-232; Madura et al., (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al., (1993) *Biotechniques* 14:920-924; Iwabuchi et al., (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 68730 or 69112 ("68730-binding proteins" or "68730-bp", or "69112-binding proteins" or "69112-bp") and are involved in 68730 or 69112 activity, respectively. Such 68730-bps and 69112-bps can be activators or inhibitors of signals by the 68730 or 69112 proteins or 68730 or 69112 targets as, for example, downstream elements of a 68730- or 69112-mediated signaling pathway.

[1456] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 68730 or 69112 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA

sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively, the 68730 or 69112 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 68730- or 69112-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 68730 or 69112 protein.

[1457] In another embodiment, modulators of 68730 or 69112 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 68730 or 69112 mRNA or protein evaluated relative to the level of expression of 68730 or 69112 mRNA or protein in the absence of the candidate compound. When expression of 68730 or 69112 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 68730 or 69112 mRNA or protein expression. Alternatively, when expression of 68730 or 69112 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 68730 or 69112 mRNA or protein expression. The level of 68730 or 69112 mRNA or protein expression can be determined by methods described herein for detecting 68730 or 69112 mRNA or protein.

[1458] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 68730 or 69112 protein can be confirmed *in vivo*, e.g., in an animal.

[1459] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 68730 or 69112 modulating agent, an antisense 68730 or 69112 nucleic acid molecule, a 68730- or 69112-specific antibody, or a 68730- or 69112-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore,

novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

[1460] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 68730 or 69112 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[1461] The 68730 and 69112 nucleotide sequences, or portions thereof, can be used to map the location of the 68730 and 69112 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 68730 and 69112 sequences with genes associated with disease.

[1462] Briefly, 68730 and 69112 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 68730 and 69112 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 68730 or 69112 sequences will yield an amplified fragment.

[1463] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al., (1983) *Science* 220:919-924).

[1464] Other mapping strategies e.g., in situ hybridization (described in Fan, Y. et al., (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 68730 or 69112 to a chromosomal location.

[1465] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique

chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

[1466] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[1467] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al., (1987) *Nature*, 325:783-787.

[1468] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 68730 or 69112 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[1469] 68730 and 69112 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification.

The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[1470] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 68730 and 69112 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[1471] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:32 and 35 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:34 or 37, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[1472] If a panel of reagents from 68730 or 69112 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 68730 and 69112 Sequences in Forensic Biology

[1473] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[1474] The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:32 or 35 (e.g., fragments derived from the noncoding regions of SEQ ID NO:32 or 35 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[1475] The 68730 and 69112 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., a tissue containing protein kinase activity. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 68730 and 69112 probes can be used to identify tissue by species and/or by organ type.

[1476] In a similar fashion, these reagents, e.g., 68730 and 69112 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

[1477] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[1478] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 68730 or 69112. The method includes one or more of the following: i) detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 68730 or 69112 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region; ii) detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 68730 or 69112 gene; iii) detecting, in a tissue of the subject, the misexpression of the 68730 or 69112 gene, at the mRNA level, e.g., detecting a non-wild type level of a mRNA; or iv)

detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, *e.g.*, detecting a non-wild type level of a 68730 or 69112 polypeptide.

[1479] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 68730 or 69112 gene; an insertion of one or more nucleotides into the gene, a point mutation, *e.g.*, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, *e.g.*, a translocation, inversion, or deletion.

[1480] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:32 or 35 naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the 68730 or 69112 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, *e.g.*, *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[1481] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 68730 or 69112 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 68730 or 69112.

[1482] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[1483] In preferred embodiments the method includes determining the structure of a 68730 or a 69112 gene, an abnormal structure being indicative of risk for the disorder.

[1484] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 68730 or 69112 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

[1485] The presence, level, or absence of 68730 and 69112 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 68730 or 69112 protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes 68730 or 69112 protein such that the presence of 68730 or 69112 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological

fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 68730 or 69112 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 68730 or 69112 genes; measuring the amount of protein encoded by the 68730 or 69112 genes; or measuring the activity of the protein encoded by the 68730 or 69112 genes.

[1486] The level of mRNA corresponding to the 68730 or 69112 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[1487] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 68730 or 69112 nucleic acid, such as the nucleic acid of SEQ ID NO:32, 34, 35 or 37, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 68730 or 69112 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

[1488] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 68730 or 69112 genes.

[1489] The level of mRNA in a sample that is transcribed by 68730 or 69112 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis, 1987, U.S. Patent No. 4,683,202), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi et al., U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques

known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[1490] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 68730 or 69112 gene being analyzed.

[1491] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 68730 or 69112 mRNA, or genomic DNA, and comparing the presence of 68730 or 69112 mRNA or genomic DNA in the control sample with the presence of 68730 or 69112 mRNA or genomic DNA in the test sample.

[1492] A variety of methods can be used to determine the level of protein encoded by 68730 or 69112. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[1493] The detection methods can be used to detect 68730 or 69112 proteins in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 68730 or 69112 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 68730 and 69112 proteins include introducing into a subject a labeled anti-68730 or anti-69112 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[1494] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 68730 or 69112 protein, and comparing the presence of 68730 or 69112 protein in the control sample with the presence of 68730 or 69112 protein in the test sample.

[1495] The invention also includes kits for detecting the presence of 68730 or 69112 in a biological sample. For example, the kit can include a compound or agent capable of detecting 68730 or 69112 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 68730 or 69112 protein or nucleic acid.

[1496] For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[1497] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein-stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[1498] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 68730 or 69112 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

[1499] In one embodiment, a disease or disorder associated with aberrant or unwanted 68730 or 69112 expression or activity is identified. A test sample is obtained from a subject and 68730 or 69112 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 68730 or 69112 protein or

nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 68730 or 69112 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

[1500] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 68730 or 69112 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cellular growth related disorder.

[1501] The methods of the invention can also be used to detect genetic alterations in 68730 and 69112 genes, thereby determining if a subject with an altered 68730 or 69112 gene is at risk for a disorder characterized by misregulation in 68730 or 69112 protein activity or nucleic acid expression, such as a cellular growth related disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 68730 or 69112 protein, or the mis-expression of the 68730 or 69112 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 68730 or 69112 gene; 2) an addition of one or more nucleotides to a 68730 or a 69112 gene; 3) a substitution of one or more nucleotides of a 68730 or 69112 gene, 4) a chromosomal rearrangement of a 68730 or 69112 gene; 5) an alteration in the level of a messenger RNA transcript of a 68730 or 69112 gene, 6) aberrant modification of a 68730 or 69112 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 68730 or 69112 gene, 8) a non-wild type level of a 68730 or 69112 protein, 9) allelic loss of a 68730 or a 69112 gene, and 10) inappropriate post-translational modification of a 68730 or a 69112 protein.

[1502] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 68730 and 69112 genes. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to

a 68730 or a 69112 gene under conditions such that hybridization and amplification of the 68730 or 69112 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

[1503] Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. et al., (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al., (1988) *Bio-Technology* 6:1197), or other nucleic acid amplification methods, followed by the detection of the amplified molecules using techniques known to those of skill in the art.

[1504] In another embodiment, mutations in a 68730 or a 69112 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis, and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[1505] In other embodiments, genetic mutations in 68730 and 69112 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two-dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. et al., (1996) *Human Mutation* 7: 244-255; Kozal, M.J. et al., (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in 68730 or 69112 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of

specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[1506] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 68730 and 69112 genes and detect mutations by comparing the sequence of the sample 68730 or 69112 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

[1507] Other methods for detecting mutations in the 68730 and 69112 genes include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al., (1985) *Science* 230:1242; Cotton et al., (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al., (1992) *Methods Enzymol.* 217:286-295).

[1508] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 68730 or 69112 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al., (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[1509] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 68730 and 69112 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al., (1989) *Proc. Natl. Acad. Sci. USA*: 86:2766, see also Cotton, (1993) *Mutat. Res.* 285:125-144; and Hayashi, (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 68730 or 69112 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes

heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al., (1991) *Trends Genet.* 7:5).

[1510] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al., (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner, (1987) *Biophys. Chem.* 265:12753).

[1511] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki et al., (1986) *Nature* 324:163); Saiki et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:6230).

[1512] Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al., (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner, (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al., (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany, (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[1513] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 68730 or a 69112 gene.

Use of 68730 and 69112 Molecules as Surrogate Markers

[1514] The 68730 and 69112 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 68730 and 69112 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 68730 or 69112 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a “surrogate marker” is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[1515] The 68730 and 69112 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a “pharmacodynamic marker” is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or

quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (*e.g.*, a 68730 or a 69112 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-68730 and anti-69112 antibodies may be employed in an immune-based detection system for a 68730 or a 69112 protein marker, or 68730- or 69112-specific radiolabeled probes may be used to detect a 68730 or 69112 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[1516] The 68730 and 69112 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a “pharmacogenomic marker” is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, *e.g.*, McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (*e.g.*, 68730 or 69112 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 68730 or 69112 DNA may correlate 68730 or 69112 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

[1517] The nucleic acid and polypeptides, fragments thereof, as well as anti-68730 and anti-69112 antibodies, and modulators of 68730 or 69112 expression or activity identified in screening assays described above (e.g., peptidomimetics, peptoids, and small molecules) (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[1518] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[1519] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the

like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[1520] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[1521] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[1522] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[1523] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[1524] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[1525] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[1526] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[1527] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that

targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1528] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[1529] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[1530] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is

described by Cruikshank et al., ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[1531] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[1532] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[1533] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, teniposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine,

lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[1534] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[1535] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[1536] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al., (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[1537] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment

[1538] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 68730 or 69112 expression or activity. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 68730 or 69112 molecules of the present invention or 68730 or 69112 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

[1539] "Treatment", as used herein, is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, palliate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[1540] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 68730 or 69112 expression or activity, by administering to the subject a 68730 or 69112 or an agent which modulates 68730 or 69112 expression or at least one 68730 or 69112 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 68730 or 69112 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 68730 or 69112 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 68730 or 69112 aberrance, for example, a 68730 or 69112 agonist

or 68730 or 69112 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[1541] It is possible that some 68730 or 69112 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[1542] As discussed, successful treatment of 68730 or 69112 disorders can be brought about by techniques that serve to inhibit the expression or activity of 68730 or 69112 target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 68730 or 69112 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')₂ and FAb expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[1543] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[1544] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[1545] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 68730 or 69112 expression is through the use of

aptamer molecules specific for 68730 or 69112 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, et al., *Curr. Opin. Chem. Biol.* 1997, 1(1): 5-9; and Patel, D.J., *Curr. Opin. Chem. Biol.* 1997 Jun;1(1):32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 68730 or 69112 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[1546] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 68730 or 69112 disorders. For a description of antibodies, see the Antibody section above.

[1547] In circumstances wherein injection of an animal or a human subject with a 68730 or 69112 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 68730 or 69112 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D., *Ann. Med.* 1999;31(1):66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A., *Cancer Treat. Res.* 1998;94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 68730 or 69112 protein. Vaccines directed to a disease characterized by 68730 or 69112 expression may also be generated in this fashion.

[1548] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco et al., (1993, *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[1549] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent,

treat or ameliorate 68730 or 69112 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

[1550] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1551] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[1552] Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 68730 or 69112 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. et al., (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J., (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to

ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. et al., (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the “free” concentration of compound which modulates the expression or activity of 68730 or 69112 can be readily monitored and used in calculations of IC_{50} .

[1553] Such “imprinted” affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . A rudimentary example of such a “biosensor” is discussed in Kriz, D. et al., (1995) *Analytical Chemistry* 67:2142-2144.

[1554] Another aspect of the invention pertains to methods of modulating 68730 and 69112 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with 68730 or 69112 or agent that modulates one or more of the activities of 68730 or 69112 protein activity associated with the cell. An agent that modulates 68730 or 69112 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 68730 or 69112 protein (e.g., a 68730 or 69112 substrate or receptor), a anti-68730 or anti-69112 antibody, a 68730 or 69112 agonist or antagonist, a peptidomimetic of a 68730 or 69112 agonist or antagonist, or other small molecule.

[1555] In one embodiment, the agent stimulates one or 68730 or 69112 activities. Examples of such stimulatory agents include active 68730 and 69112 proteins and nucleic acid molecules encoding 68730 or 69112. In another embodiment, the agent inhibits one or more 68730 or 69112 activities. Examples of such inhibitory agents include antisense 68730 and 69112 nucleic acid molecules, anti-68730 and anti-69112 antibodies, and 68730 and 69112 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 68730 or a 69112 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 68730 or 69112 expression or activity. In another embodiment, the method

involves administering a 68730 or 69112 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 68730 or 69112b expression or activity.

[1556] Stimulation of 68730 or 69112 activity is desirable in situations in which 68730 or 69112 is abnormally downregulated and/or in which increased 68730 or 69112 activity is likely to have a beneficial effect. For example, stimulation of 68730 or 69112 activity is desirable in situations in which a 68730 or a 69112 is downregulated and/or in which increased 68730 or 69112 activity is likely to have a beneficial effect. Likewise, inhibition of 68730 or 69112 activity is desirable in situations in which 68730 or 69112 is abnormally upregulated and/or in which decreased 68730 or 69112 activity is likely to have a beneficial effect.

[1557] The 68730 and 69112 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of cellular proliferative and/or differentiative disorders, cardiovascular disorders, as described above, as well as disorders associated with bone metabolism, hematopoietic disorders, liver disorders, viral diseases, pain or metabolic disorders.

[1558] Aberrant expression and/or activity of 68730 or 69112 molecules may mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which may ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 68730 or 69112 molecules effects in bone cells, e.g. osteoclasts and osteoblasts, that may in turn result in bone formation and degeneration. For example, 68730 or 69112 molecules may support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 68730 or 69112 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus may be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[1559] Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, α 1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein may be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyl dopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[1560] Additionally, 68730 and 69112 molecules may play an important role in the etiology of certain viral diseases, including but not limited to, Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 68730 or 69112 activity can be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 68730 and 69112 modulators can be used in the treatment and/or diagnosis of virus-associated carcinomas, especially hepatocellular cancers.

[1561] Additionally, 68730 or 69112 may play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, bulimia, cachexia, lipid disorders, and diabetes. Examples of

pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L., (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; or chest pain.

Pharmacogenomics

[1562] The 68730 and 69112 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 68730 or 69112 activity (e.g., 68730 or 69112 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 68730 or 69112 associated disorders (e.g., cellular growth related disorders) associated with aberrant or unwanted 68730 or 69112 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 68730 or a 69112 molecule or a 68730 or a 69112 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 68730 or 69112 molecule or modulator.

[1563] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. et al. (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11) :983-985 and Linder, M.W. et al. (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[1564] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high-resolution map can be generated from a combination of some ten million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[1565] Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 68730 or a 69112 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[1566] Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 68730 or 69112 molecule or 68730 or 69112 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[1567] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 68730 or 69112

molecule or 68730 or 69112 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[1568] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 68730 or 69112 genes of the present invention, wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 68730 or 69112 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., cancer cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

[1569] Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 68730 or 69112 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 68730 or 69112 gene expression, protein levels, or upregulate 68730 or 69112 activity, can be monitored in clinical trials of subjects exhibiting decreased 68730 or 69112 gene expression, protein levels, or downregulated 68730 or 69112 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 68730 or 69112 gene expression, protein levels, or downregulate 68730 or 69112 activity, can be monitored in clinical trials of subjects exhibiting increased 68730 or 69112 gene expression, protein levels, or upregulated 68730 or 69112 activity. In such clinical trials, the expression or activity of a 68730 or 69112 gene, and preferably, other genes that have been implicated in, for example, a 68730- or 69112-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

[1570] In another aspect, the invention features, a method of analyzing a plurality of capture probes. The method can be used, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, wherein each address of the plurality is positionally distinguishable from every other address of the plurality, and each address of the plurality has a unique capture probe, e.g., a nucleic acid or peptide sequence; contacting the array with a 68730 or 69112 molecule, e.g., a 68730 or 69112 nucleic acid, preferably purified, a 68730 or 69112 polypeptide, preferably purified, or an anti-68730 or anti-69112 antibody, preferably purified, and evaluating the plurality of capture probes. Binding with a capture probe at an address of the plurality, is detected, e.g., by a signal generated from a label attached to the 68730 or 69112 nucleic acid, polypeptide, or antibody.

[1571] The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

[1572] The method can include contacting the 68730 or 69112 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

[1573] The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 68730 or 69112. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

[1574] The method can be used to detect SNPs, as described above.

[1575] In another aspect, the invention features, a method of analyzing a plurality of probes. The method is useful, e.g., for analyzing gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality having a

unique capture probe, e.g., wherein the capture probes are from a cell or subject which expresses or misexpresses 68730 or 69112 or from a cell or subject in which a 68730 or 69112 mediated response has been elicited, e.g., by contact of the cell with 68730 or 69112 nucleic acid or protein, or administration to the cell or subject 68730 or 69112 nucleic acid or protein; contacting the array with one or more inquiry probe, wherein an inquiry probe can be a nucleic acid, polypeptide, or antibody (which is preferably other than 68730 or 69112 nucleic acid, polypeptide, or antibody); providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., wherein the capture probes are from a cell or subject which does not express 68730 or 69112 (or does not express as highly as in the case of the 68730 or 69112 positive plurality of capture probes) or from a cell or subject which in which a 68730- or 69112-mediated response has not been elicited (or has been elicited to a lesser extent than in the first sample); contacting the array with one or more inquiry probes (which is preferably other than a 68730 or 69112 nucleic acid, polypeptide, or antibody), and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by signal generated from a label attached to the nucleic acid, polypeptide, or antibody.

[1576] In another aspect, the invention features, a method of analyzing 68730 or 69112, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 68730 or 69112 nucleic acid or amino acid sequence; comparing the 68730 or 69112 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 68730 or 69112.

[1577] Preferred databases include GenBank™. The method can include evaluating the sequence identity between a 68730 or 69112 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet.

[1578] In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 68730 or 69112. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with different labels, such that

an oligonucleotides which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

[1579] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

Equivalents

[1580] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

VI. 52908, A HUMAN POTASSIUM CHANNEL, AND USES THEREOF

BACKGROUND OF THE INVENTION

[1581] Potassium (K^+) channels are ubiquitous proteins which are involved in the setting of the resting membrane potential as well as in the modulation of the electrical activity of cells. In excitable cells, K^+ channels influence action potential waveforms, firing frequency, and neurotransmitter secretion (Rudy, B. (1988) *Neuroscience*, 25, 729-749; Hille, B. (1992) *Ionic Channels of Excitable Membranes*, 2nd Ed.). In non-excitable cells, they are involved in hormone secretion, cell volume regulation and potentially in cell proliferation and differentiation (Lewis *et al.* (1995) *Annu. Rev. Immunol.*, 13, 623-653). Developments in electrophysiology have allowed the identification and the characterization of an astonishing variety of K^+ channels that differ in their biophysical properties, pharmacology, regulation and tissue distribution (Rudy, B. (1988) *Neuroscience*, 25, 729-749; Hille, B. (1992) *Ionic Channels of Excitable Membranes*, 2nd Ed.). More recently, cloning efforts have shed considerable light on the mechanisms that determine this functional diversity. Furthermore, analyses of structure-function relationships have provided an important set of data concerning the molecular basis of the biophysical properties (selectivity, gating, assembly) and the pharmacological properties of cloned K^+ channels.

[1582] Potassium channels are potassium-ion selective, and can determine membrane excitability (the ability of, for example, a neuron to respond to a stimulus and

convert it into an impulse). Potassium channels can also influence the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation. Potassium channels are typically expressed in electrically excitable cells, *e.g.*, neurons, muscle, endocrine, and egg cells, and may form heteromultimeric structures, *e.g.*, composed of pore-forming α and cytoplasmic β subunits. Potassium channels may also be found in non-excitable cells, where they may play a role in, *e.g.*, signal transduction. Examples of potassium channels include: (1) the voltage-gated potassium channels, (2) the ligand-gated potassium channels, *e.g.*, neurotransmitter-gated potassium channels, and (3) cyclic-nucleotide-gated potassium channels. Voltage-gated and ligand-gated potassium channels are expressed in the brain, *e.g.*, in brainstem monoaminergic and forebrain cholinergic neurons, where they are involved in the release of neurotransmitters, or in the dendrites of hippocampal and neocortical pyramidal cells, where they are involved in the processes of learning and memory formation.

[1583] Voltage-gated potassium channels (Kv) include: (1) the delayed potassium channels, which repolarize the membrane after each action potential to prepare the cell to fire again; (2) the early potassium channels, which open when the membrane is depolarized and act to reduce the rate of firing at levels of stimulation which are just above the threshold required for firing; and (3) the calcium-activated potassium channels, which act along with the voltage-gated calcium channels to decrease the response of the cell to an unchanging prolonged stimulation, a process called adaptation. In addition to being critical for action potential conduction, the voltage-gated potassium channels also play a role in neurotransmitter release. As a result of these activities, voltage-gated potassium channels are important in controlling neuronal excitability (Hille B., *Ionic Channels of Excitable Membranes*, Second Edition, Sunderland, MA: Sinauer, (1992)).

[1584] There is a surprising amount of structural and functional diversity within the voltage-gated potassium channels. This diversity is generated both by the existence of multiple genes and by alternative splicing of RNA transcripts produced from the same gene. Nonetheless, the amino acid sequences of the known voltage-gated potassium channels show similarity. The *Drosophila* SH locus was the first potassium channel structural gene to be isolated (Kamb A. et al. (1987) *Cell* 50: 405). Since then, a number of additional potassium channel genes have been cloned from *Drosophila* and other organisms (Baumann A. et al. (1988) *EMBO J.* 7: 2457). One of these genes is the X-linked *EAG* locus, which was originally identified in *Drosophila* on the basis of mutations that cause a leg-shaking

phenotype (Kaplan W.D. et al. (1969) *Genetics* 61: 399). Electrophysiological studies revealed that *EAG* mutations caused spontaneous repetitive firing in motor axons and elevated transmitter release at the larval neuromuscular junction (Ganetzky B. et al. (1985) *Trends Neurosci.* 8:322). The striking hyperexcitability of *EAG* mutants demonstrates the importance of *EAG* channels in maintaining normal neuronal excitability in *Drosophila* (Ganetzky B. et al. (1983) *J. Neurogenet.* 1: 17-28).

[1585] *EAG*, along with m-*EAG*, *ELK*, and h-*ERG* (*EAG*-related gene), define a family of potassium channel genes in *Drosophila* and mammals. A distinctive feature of the *EAG/ERG* family is the homology to cyclic nucleotide binding domains of cyclic nucleotide-gated cation channels and cyclic nucleotide-activated protein kinases (Kaupp, U.B. et al. (1991) *Trends Neurosci.* 14: 150-157). However, unlike the vertebrate cyclic nucleotide-gated cation channels, which are relatively voltage-insensitive, activation of *EAG/ERG* channels shows a very steep voltage dependence (Robertson, G. et al. (1993) *Biophys. J.* 64: 430). In addition, whereas cyclic nucleotide-activated cation channels show little selectivity among monovalent and divalent cations, *EAG* is strongly selective for K^+ over Na^+ . The *EAG/ERG* family may thus be an evolutionary link between voltage-activated potassium channels and cyclic nucleotide-gated cation channels with intermediate structural and functional properties.

[1586] Functional diversity of K^+ channels also arises from the existence of a great number of genes coding for pore-forming subunits, as well as for other associated regulatory subunits. Two main structural families of pore-forming subunits have been identified, one of which consists of subunits with a conserved hydrophobic core containing six transmembrane domains (TMDs), and the other have only two TMDs. The K^+ channel α subunits of the six TMD family participate in the formation of outward rectifier voltage-gated (K_v) and Ca^{2+} -dependent K^+ channels. The fourth TMD contains repeated positive charges involved in the voltage gating of these channels and hence in their outward rectification (Logothetis et al. (1992) *Neuron*, 8, 531-540; Bezanilla et al. (1994) *Biophys. J.* 66, 1011-1021).

[1587] In both six TMD and two TMD pore-forming subunit families, different subunits coded by different genes can associate to form heterotetramers with new channel properties (Isacoff et al., (1990) *Nature*, 345, 530-534). A selective formation of heteropolymeric channels may allow each cell to develop the best K^+ current repertoire

suites to its function. Pore-forming α subunits of Kv channels are classified into different subfamilies according to their sequence similarity (Chandy *et al.* (1993) *Trends Pharmacol. Sci.*, 14: 434). Tetramerization is believed to occur preferentially between members of each subgroup (Covarrubias *et al.* (1991) *Neuron*, 7, 763-773). The domain responsible for this selective association is localized in the N-terminal region and is conserved between members of the same subgroup. This domain is necessary for hetero- but not homo-multimeric assembly within a subfamily and prevents co-assembly between subfamilies.

SUMMARY OF THE INVENTION

[1588] The present invention is based, in part, on the discovery of a novel potassium channel family members, referred to herein as "52908." The nucleotide sequence of the cDNA encoding 52908 is shown in SEQ ID NO:47, and the amino acid sequence of the 52908 polypeptide is shown in SEQ ID NO:48. In addition, the nucleotide sequence of the coding region of 52908 is depicted in SEQ ID NO:49.

[1589] Accordingly, in one aspect, the invention features nucleic acid molecules that encode a 52908 protein or polypeptide, e.g., a biologically active portion of the 52908 protein. In a preferred embodiment the isolated nucleic acid molecules encode polypeptides having the amino acid sequence SEQ ID NO:48. In other embodiments, the invention provides isolated 52908 nucleic acid molecules having the nucleotide sequences of one of SEQ ID NO:47, and SEQ ID NO:49.

[1590] In still other embodiments, the invention provides nucleic acid molecules that have sequences that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequences of one of SEQ ID NO:47, and SEQ ID NO:49. In other embodiments, the invention provides nucleic acid molecules which hybridize under stringent hybridization conditions with a nucleic acid molecule having a sequence comprising the nucleotide sequence of one of SEQ ID NO:47, and SEQ ID NO:49, wherein the nucleic acids encode full length 52908 protein or an active fragment thereof.

[1591] In a related aspect, the invention further provides nucleic acid constructs that include 52908 nucleic acid molecules described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included are vectors and host cells containing the 52908 nucleic acid molecules of the invention, e.g., vectors and host cells suitable for producing 52908 polypeptides.

[1592] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for detection of 52908-encoding nucleic acids.

[1593] In still another related aspect, isolated nucleic acid molecules that are antisense to 52908-encoding nucleic acid molecules are provided.

[1594] In another aspect, the invention features 52908 polypeptides, and biologically active or antigenic fragments thereof, that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of 52908-mediated or 52908-related disorders.

[1595] In another embodiment, the invention provides 52908 polypeptides having a 52908 activity. Preferred polypeptides are 52908 proteins including at least one conserved 52908 domain, e.g., a potassium channel domain (e.g., a pore loop domain); and 52908 proteins including one or more transmembrane domains.

[1596] In other embodiments, the invention provides 52908 polypeptides, e.g., 52908 polypeptides having the amino acid sequences shown in SEQ ID NO:48; amino acid sequences that are substantially identical to the amino acid sequences shown in SEQ ID NO:48; or amino acid sequences encoded by nucleic acid molecules having nucleotide sequences which hybridize under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:47, SEQ ID NO:49, wherein the nucleic acids encode full length 52908 protein or an active fragment thereof.

[1597] In a related aspect, the invention further provides nucleic acid constructs that include 52908 nucleic acid molecules described herein.

[1598] In a related aspect, the invention provides 52908 polypeptides or fragments operatively linked to non-52908 polypeptides to form fusion proteins.

[1599] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably, specifically or selectively bind, 52908 polypeptides.

[1600] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 52908 polypeptides or nucleic acids.

[1601] In still another aspect, the invention provides a process for modulating 52908 polypeptide or nucleic acid expression or activity, e.g., using the compounds identified in the screens described herein. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 52908 polypeptides or nucleic acids, such as CNS-related (e.g., neurological) disorders; pain and metabolic disorders;

cardiovascular disorders; cellular proliferative, growth, differentiative, and/or migration disorders; immune, e.g., inflammatory, disorders; disorders associated with bone metabolism; endothelial cell disorders; liver disorders; and viral diseases.

[1602] The invention also provides assays for determining the activity of or the presence or absence of 52908 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[1603] In further aspect the invention provides assays for determining the presence or absence of a genetic alteration in a 52908 polypeptide or nucleic acid molecule, including for disease diagnosis.

[1604] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

DETAILED DESCRIPTION OF THE INVENTION

[1605] The human 52908 cDNA sequence (SEQ ID NO:47), which is approximately 3164 nucleotide residues long including non-translated regions, contains a methionine-initiated coding sequence of about 2877 nucleotide residues, excluding termination codon (i.e., nucleotide residues 1-2877 of SEQ ID NO:47; also shown in SEQ ID NO:49). The coding sequence encodes a 958 amino acid protein having the amino acid sequence SEQ ID NO:48.

[1606] Human 52908 contain the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997, Protein 28:405-420): transmembrane domains at about amino acid residues 115-131 of SEQ ID NO:48, and at about 261-282, 295-317, 345-361, 371-390, 398-420, and 491-515 of SEQ ID NO:48; and post translational modification sites including: predicted N-glycosylation sites (Pfam accession number PS00001) at about amino acid residues 271-274 of SEQ ID NO:48, at about amino acid residues 97-99, 104-106, 121-123, and 226-228 of SEQ ID NO:55, and at about amino acid residues 218-221, 449-452, 510-513, and 731-734 of SEQ ID NO:48; predicted protein kinase C phosphorylation sites (Pfam accession number PS00005) at about amino acid residues 26-28, 59-61, 214-216, 233-235, 341-343, 364-366, 404-406, and 513-515 of SEQ ID NO:48, and at about amino acid residues 26-28, 105-107, 140-142, 145-147, 170-172, 220-222, 288-290, 377-379, 488-490, 522-524, 861-863, 868-870, and 878-880 of SEQ ID NO:48; predicted casein kinase II phosphorylation sites (Pfam accession number

PS00006) located at about amino acid residues 32-35, 44-47, 77-80, 82-85, and 433-436 of SEQ ID NO:48, at about amino acid residues 129-132 and 195-198 of SEQ ID NO:55, and at about amino acid residues 13-16, 55-58, 200-203, 283-286, 301-304, 326-239, 363-366, 458-461, 486-489, 670-673, 678-681, 706-709, 740-743, 763-766, 777-780, 853-856, 886-889, and 918-921 of SEQ ID NO:48; predicted N-myristoylation sites (Pfam accession number PS00008) at about amino acid residues 131-136 and 490-495 of SEQ ID NO:48, at about amino acid residues 135-140 and 142-147 of SEQ ID NO:55, and at about amino acid residues 88-93, 155-160, 169-174, 364-369, 373-378, 441-446, 450-455, 509-514, 562-567, 597-602, and 936-941 of SEQ ID NO:48; and a predicted amidation site (Pfam accession number PS00009) at about amino acid residues 4-7 of SEQ ID NO:48 and at about amino acid residues 444-447 of SEQ ID NO:48.

[1607] Additionally, 52908 contains the following regions or other structural features: a pore loop domain at about amino acid residues 463-482 of SEQ ID NO:48; a PAS domain at about amino acid residues 41-59 of SEQ ID NO:48; a PAC domain at about amino acid residues 93-135 of SEQ ID NO:48; a cyclic nucleotide gated channel transmembrane region at about amino acid residues 608-699 of SEQ ID NO:48; and a cyclic nucleotide gated channel transmembrane region at about amino acid residues 341-580 of SEQ ID NO:48.

[1608] A hydropathy plot of human 52908 can be performed. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, e.g., the sequences of about residues 295-317, 345-361, and 491-515 of SEQ ID NO:48; all or part of a hydrophilic sequence, e.g., the sequences of residues 318-344, 533-550, and 752-770 of SEQ ID NO:48; a sequence which includes a cysteine residue; or a glycosylation site.

Table 3: Summary of Sequence Information for 52908

Gene	cDNA	ORF	Polypeptide
52908	SEQ ID NO:47	SEQ ID NO:49	SEQ ID NO:48

[1609] The 52908 protein contains a number of structural characteristics in common with members of the potassium channel family. The term “family” when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having common structural domains (e.g., a transmembrane domain, a pore

loop domain) or motifs and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologues of non-human origin, e.g., potassium channel proteins for any species described in the art (e.g., Steiner et al. (1995) *Mol. Microbiol.* 16:825-834, and references cited therein). Members of a family can also have common functional characteristics.

[1610] In one embodiment, the 52908 protein are members of the potassium channel family. In particular, the 52908 protein resembles a member of the *erg* (EAG-related gene) family of potassium channels, and as described herein.

[1611] As used herein, a “potassium channel” includes a protein or polypeptide which is involved in receiving, conducting, and transmitting signals in an electrically excitable cell, e.g., a neuronal cell or a muscle cell. Potassium channels are potassium ion selective, and can determine membrane excitability (the ability of, for example, a neuron to respond to a stimulus and convert it into an impulse). Potassium channels can also influence the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation. Examples of potassium channels include: (1) the voltage-gated potassium channels, (2) the ligand-gated potassium channels, e.g., neurotransmitter-gated potassium channels, and (3) cyclic-nucleotide-gated potassium channels. Voltage-gated and ligand-gated potassium channels are expressed in the brain, e.g., in brainstem monoaminergic and forebrain cholinergic neurons, where they are involved in the release of neurotransmitters, or in the dendrites of hippocampal and neocortical pyramidal cells, where they are involved in the processes of learning and memory formation. For a detailed description of potassium channels, see Kandel E.R. *et al.*, Principles of Neural Science, second edition, (Elsevier Science Publishing Co., Inc., N.Y. (1985)), the contents of which are incorporated herein by reference.

[1612] The 52908 protein, fragments thereof, and derivatives, and other variants of the sequences in SEQ ID NO:48 thereof are collectively referred to as “polypeptides or proteins of the invention” or “52908 polypeptides or proteins”. Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as “nucleic acids of the invention” or “52908 nucleic acids.”

[1613] To determine whether a polypeptide or protein of interest has a conserved sequence or domain common to members of a protein family, the amino acid sequence of the protein can be searched against a database of profile hidden Markov models (profile HMMs), which uses statistical descriptions of a sequence family's consensus (e.g., HMMER, version 2.1.1) and PFAM, a collection of multiple sequence alignments and hidden Markov models covering many common protein domains (e.g., PFAM, version 5.5) using the default parameters. For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (e.g., to 8 bits). A description of the PFAM database can be found in Sonnhammer et al., (1997) *Proteins* 28(3):405-420 and a detailed description of HMMs can be found, for example, in Gribskov et al., (1990) *Meth. Enzymol.* 183:146-159; Gribskov et al., (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh et al., (1994) *J. Mol. Biol.* 235:1501-1531; and Stultz et al., (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. See also, for example, The HMMER User's Guide available online. For general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer et al. (1997) *Protein* 28:405-420.

[1614] Using such search tools, a pore loop domain profile was identified in the amino acid sequences of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a pore loop domain of human 52908 are within the scope of the invention. As used herein, the term "Pore loop" or "P-loop" includes an amino acid sequence of about 10-45 amino acid residues in length, preferably about 15-30 amino acid residues in length, and most preferably about 19-20 amino acid residues in length, which is hydrophobic and which is involved in lining the potassium channel pore. A P-loop is typically found between transmembrane domains of potassium channels and is believed to be a major determinant of ion selectivity in potassium channels. Preferably, P-loops contain a "potassium channel pore motif", which has the sequence of G-[HYDROPHOBIC AMINO ACID]-G sequence, e.g., a GYG, GLG, or GFG sequence. P-loops are described in, for example, Warmke et al. (1991) *Science* 252:1560-1562; Zagotta W.N. et al., (1996) *Annual Rev. Neurosci.* 19:235-63; Pongs, O. (1993) *J. Membr. Biol.*, 136:1-8; Heginbotham et al. (1994) *Biophys. J.* 66:1061-1067; Mackinnon, R. (1995) *Neuron* 14:889-892; Pascual et al., (1995) *Neuron*,

14:1055-1063), the contents of which are incorporated herein by reference. Amino acid residues 463-482 of SEQ ID NO:48 comprise a P-loop domain, and the potassium channel pore motif is located from amino acid residues 478-480 of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a P-loop domain of human 52908 is within the scope of the invention.

[1615] A PAS domain profile was identified in the amino acid sequence of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a PAS domain of human 52908 is within the scope of the invention.

[1616] A 52908 polypeptide can further include a "PAS domain" or regions homologous with a "PAS domain". As used herein, the term "PAS domain" includes an amino acid sequence of about 5-50, preferably 10-40, more preferably 10-30 amino acid residues, and which is involved in ligand and/or protein-protein interactions. Preferably, the PAS domain interacts with the body of the channel, affecting gating, inactivation, and/or voltage sensitivity. PAS domains, and the PAC sequences which usually accompany them are found in many diverse organisms, are often found in circadian-cycle regulatory proteins, and have sensory and protein-binding functions. Preferably, the PAS domain is located at the N-terminal cytoplasmic region of the 52908 polypeptide. The PAS domain has been assigned the PFAM Accession PF00989 (SEQ ID NO:50). The single occurrence for this domain in 52908 as predicted by PFAM has a bit score of 7.5 and an E-value of 1.9.

[1617] Amino acid residues 41-59 of SEQ ID NO:48 comprise a PAS domain, which can have the following sequence: I-Xaa-Y-Xaa-N-Xaa(4)-E-L-T-G-L-S-R-Xaa-E-V (SEQ ID NO:57).

[1618] A PAC domain was identified in the amino acid sequence of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a cyclic nucleotide gated channel transmembrane region of human 52908 is within the scope of the invention. A PAC domain was identified both through searching against a database of profile hidden Markov models (profile HMMs), and through searching for complete domains in SMART database online.

[1619] A 52908 polypeptide can further include a "PAC domain" or regions homologous with a "PAC domain". As used herein, the term "PAC domain" includes an amino acid sequence of about 30-60, preferably about 35-50, more preferably about 40-45

amino acid residues in length; which has a bit score for the alignment of the sequence to the PAC domain sequence (HMM) of at least 10, more preferably at least 20; and which has an E-value of 1 or less, preferably .001 or less, preferably .0001 or less, still more preferably $1.8e-100$ or less. PAC domains are C-terminal to a subset of all known PAS domains, and are proposed to contribute to the PAS domain characteristic fold. The PAC domain has been assigned the PFAM Accession PF00785 (SEQ ID NO:51). The single occurrence for this domain in 52908 as predicted by PFAM has a bit score of 23.3 and an E-value of $5.3e-05$.

[1620] In one embodiment, the PAC domain can have the following consensus sequence: R-K-D-Xaa-S-Xaa(4)-L-V-Xaa(3)-P-Xaa(3)-E-D-G-Xaa-V-Xaa(8)-D (SEQ ID NO:58).

[1621] Amino acid residues 93-135 of SEQ ID NO:48 comprise a PAC domain, and the PAC domain consensus sequence is located from 100-131 of SEQ ID NO:48.

[1622] A cyclic nucleotide gated channel transmembrane region profile was identified in the amino acid sequence of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a cyclic nucleotide gated channel transmembrane region of human 52908 is within the scope of the invention.

[1623] A 52908 polypeptide can further include a “cyclic nucleotide gated channel transmembrane region” or regions homologous with a “cyclic nucleotide gated channel transmembrane region.” As used herein, the term “cyclic nucleotide gated channel transmembrane region” includes an amino acid sequence of about 100-400, preferably about 150-300, more preferably about 200-250 amino acid residues in length; which has a bit score for the alignment of the sequence to the cyclic nucleotide gated channel transmembrane region (HMM) of at least 100, preferably at least 200, more preferably at least 300, and still more preferably at least 400; and which has an E-value of 1 or less, preferably $1.0e-10$ or less, preferably $1.0e-50$ or less, still more preferably $1.8e-100$ or less. Cyclic nucleotide gated channels are involved in odorant signal transduction in olfactory epithelium and visual signal transduction in photoreceptor cells such as rod cells. The cyclic nucleotide gated channel transmembrane region has been assigned the PFAM Accession PF00914 (SEQ ID NO:52). The single occurrence for this domain in 52908 as predicted by PFAM has a bit score of 404.8 and an E-value of $3.9e-117$.

[1624] In one embodiment, the cyclic nucleotide gated channel transmembrane region domain can have the following consensus sequence: W-F-L-[LI]-Xaa(5)-P-F-D-L-L-Xaa(4)-G-S-D-E-Xaa(n1)-L-L-Xaa(3)-R-L-L-R-L-Xaa-R-V-A-Xaa(3)-D-R (SEQ ID NO:59).

[1625] In this consensus sequence, n1 can be 1-8, preferably 2-6, and more preferably 3-5. Amino acid residues 341-580 of SEQ ID NO:48 comprise a cyclic nucleotide gated channel transmembrane region domain, and the cyclic nucleotide gated channel transmembrane region domain consensus sequence is located from 345-392 of SEQ ID NO:48.

[1626] A cyclic nucleotide binding domain (CNBD) profile was identified in the amino acid sequence of SEQ ID NO:48. Accordingly, proteins having at least 50-60% homology, preferably about 60-70%, more preferably about 70-80%, or about 80-90% homology with a cyclic nucleotide binding domain (CNBD) of human 52908 is within the scope of the invention.

[1627] A 52908 polypeptide can further include a “cyclic nucleotide binding domain (CNBD)” or regions homologous with a “CNBD”. As used herein, the term “CNBD” includes an amino acid sequence of about 50-150 amino acids, preferably about 60-130 amino acids, more preferably about 70-120 amino acids; which has a bit score for the alignment of the sequence to the CNBD domain (HMM) of at least 50, preferably at least 60, more preferably at least 75; and which has an E-value of 1 or less, preferably 1.0e-5 or less, preferably 1.0e-10 or less, still more preferably 1.0e-19 or less. CNBD domains are capable of binding a cyclic nucleotide, and are found in such proteins as prokaryotic catabolite gene activator protein (CAP), cAMP- and cGMP-depedent protein kinases (cAPK and cGPK), and vertebrate cyclic nucleotide-gated ion channels. The cyclic nucleotide binding domain (CNBD) has been assigned the PFAM Accession PF00027 (SEQ ID NO:53). The single occurrence for this domain in 52908 as predicted by PFAM has a bit score of 78.8 and an E-value of 1.1e-19.

[1628] In one embodiment the cyclic nucleotide binding domain (CNBD) can have one or more of the following consensus sequences: [LIVM]-[VIC]-Xaa(2)-G-[DENQTA]-Xaa-[GACL]-Xaa(2)-[LIVMFY](4)-Xaa(2)-G (SEQ ID NO:60) and [LIVMF]-G-E-Xaa-[GASV]-[LIVMS]-Xaa(n1)-[RS]-[STAQ]-A-Xaa-[LIVMA]-Xaa-[STACV] (SEQ ID NO:61).

[1629] In this consensus sequence, n1 can be 2-15, preferably 3-14, more preferably 5-13. The conserved glycine residues (two in the first consensus sequence, and one in the second (underlined)) are thought to be essential for the maintenance of the structural integrity of the beta-barrel structure which is characteristic of proteins possessing cyclic nucleotide binding domains.

[1630] Amino acid residues 615-687 of SEQ ID NO:48 comprise a cyclic nucleotide binding domain (CNBD), and cyclic nucleotide binding domain (CNBD) consensus sequences are located from amino acid residues 621-637 and amino acid residues 657-676 of SEQ ID NO:48.

[1631] In one embodiment, the 52908 polypeptide or protein has one or more of the following: a pore loop domain or region which includes at least about 10-45, preferably about 15-30, more preferably about 19-20 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a pore loop domain, e.g., the pore loop domain of human 52908 (e.g., residues 463-482 of SEQ ID NO:48); a PAS domain or region which includes at least about 5-50, preferably 10-40, more preferably 10-30 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 41-59 of SEQ ID NO:48); a PAC domain or region which includes at least about 30-60, preferably about 35-50, more preferably about 40-45 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 93-135 of SEQ ID NO:48); a cyclic nucleotide gated channel transmembrane region which includes at least about 100-400, preferably about 150-300, more preferably about 200-250 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide gated channel transmembrane region, e.g., the cyclic nucleotide gated channel transmembrane region domain of human 52908 (e.g., residues 341-580 of SEQ ID NO:48); and a cyclic nucleotide binding domain or region which includes at least about 50-150, preferably about 60-130, more preferably about 70-120 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide binding domain, e.g., the cyclic nucleotide binding domain of human 52908 (e.g., residues 615-687 of SEQ ID NO:48).

[1632] In another embodiment, the 52908 polypeptide or protein has one or more of the following: a pore loop domain or region which includes at least about 10-45, preferably about 15-30, more preferably about 19-20 amino acid residues and has at least about 60%,

70%, 80%, 90%, 95%, 99%, or 100% homology with a pore loop domain, e.g., the pore loop domain of human 52908 (e.g., residues 463-482 of SEQ ID NO:48); a PAS domain or region which includes at least about 5-50, preferably 10-40, more preferably 10-30 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 41-59 of SEQ ID NO:48); a PAC domain or region which includes at least about 30-60, preferably about 35-50, more preferably about 40-45 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 93-135 of SEQ ID NO:48); a cyclic nucleotide gated channel transmembrane region which includes at least about 100-400, preferably about 150-300, more preferably about 200-250 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide gated channel transmembrane region, e.g., the cyclic nucleotide gated channel transmembrane region domain of human 52908 (e.g., residues 341-580 of SEQ ID NO:48); and a cyclic nucleotide binding domain or region which includes at least about 50-150, preferably about 60-130, more preferably about 70-120 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide binding domain, e.g., the cyclic nucleotide binding domain of human 52908 (e.g., residues 615-687 of SEQ ID NO:48), and has at least one of the PAC domain, PAS domain, cyclic nucleotide gated channel transmembrane region, and cyclic nucleotide binding domain (CNBD) consensus sequences described herein.

[1633] In still another embodiment, the 52908 polypeptide or protein has one or more of the following: a pore loop domain or region which includes at least about 10-45, preferably about 15-30, more preferably about 19-20 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a pore loop domain, e.g., the pore loop domain of human 52908 (e.g., residues 463-482 of SEQ ID NO:48); a PAS domain or region which includes at least about 5-50, preferably 10-40, more preferably 10-30 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 41-59 of SEQ ID NO:48); a PAC domain or region which includes at least about 30-60, preferably about 35-50, more preferably about 40-45 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a PAS domain, e.g., the PAS domain of human 52908 (e.g., residues 93-135 of SEQ ID NO:48); a cyclic nucleotide gated channel transmembrane region which includes at least about 100-400, preferably about 150-300,

more preferably about 200-250 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide gated channel transmembrane region, e.g., the cyclic nucleotide gated channel transmembrane region domain of human 52908 (e.g., residues 341-580 of SEQ ID NO:48); and a cyclic nucleotide binding domain or region which includes at least about 50-150, preferably about 60-130, more preferably about 70-120 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cyclic nucleotide binding domain, e.g., the cyclic nucleotide binding domain of human 52908 (e.g., residues 615-687 of SEQ ID NO:48), and at least potassium channel biological activity as described herein.

[1634] In one embodiment, a 52908 protein includes one or more transmembrane domains. As used herein, the term “transmembrane domain” includes an amino acid sequence of about 5 amino acid residues in length that spans the plasma membrane. More preferably, a transmembrane domain includes about at least 10, 15, 20 or 22-25 amino acid residues and spans a membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an alpha-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, or 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, <http://pfam.wustl.edu/cgi-bin/getdesc?name=7tm-1>, and Zagotta W.N. et al. (1996, Annu. Rev. Neurosci. 19: 235-263), the contents of which are incorporated herein by reference. Transmembrane domains exist at least about amino acid residues 115-131 of SEQ ID NO:48 and about amino acid residues 261-282, 295-317, 345-361, 371-390, 398-420, and 491-515 of SEQ ID NO:48.

[1635] A 52908 family member can include one or more transmembrane domains. Furthermore, a 52908 family member can contain one or more N-glycosylation sites, protein kinase C phosphorylation sites, casein kinase II phosphorylation sites, N-myristoylation sites, and amidation sites.

[1636] Furthermore, a 52908 family member can include one or more pore loop domains, PAS domains, PAC domains, cyclic nucleotide gated channel transmembrane regions, cyclic nucleotide gated channel transmembrane regions, and/or cyclic nucleotide binding domains.

[1637] 52908 is homologous to a number of proteins belonging to the erg (human Eag-related gene) family, which is a family of voltage-gated potassium channels that are important in determining the threshold firing property of neurons. Some members of the erg

family are expressed exclusively in the nervous system (e.g., in sympathetic ganglia), and in *Drosophila*, mutations in the *erg* gene cause neurological defects (Shi et al. (1997) *Jour. of Neuro.* 17(24):9423-9432). Mutations in genes of this family also have been shown to cause QT syndrome, which gives rise to arrhythmias and an increased incidence of sudden death.

[1638] Specifically, 52908 is homologous to hERG2, a human Eag-related gene member 2 protein (Genbank accession number AAG40871; SEQ ID NO:54). An alignment of hERG2 with 52908 reveals 99.5% identity and 99.6% similarity. 52908 is also homologous to rERG3, a rat Eag-related gene member 3 protein (rERG3)(Genbank accession number AAB95841; SEQ ID NO:56). An alignment of rERG3 with 52908 reveals 55.3% identity and 59.3% similarity. [Both alignments were performed using the GAP alignment program with a BLOSUM62 scoring matrix, a gap open penalty of 12, and a gap extend penalty of 4.]

[1639] Lastly, 52908 is homologous to rERG2, a rat Eag-related gene member 2 protein (rERG2)(Genbank accession number AAB95842; SEQ ID NO:55). An alignment of rERG2 with 52908 reveals 90.0% identity and 90.6% similarity (the alignment was performed using the GAP alignment program with a BLOSUM62 scoring matrix, a gap open penalty of 12, and a gap extend penalty of 4). 6 transmembrane domains are present in both 52908 and rERG2 (from amino acids 261-282, 295-317, 345-361, 371-390, 398-420, and 491-515 of SEQ ID NO:48, as predicted by HMMer, as described herein; and from amino acids 262-282, 301-324, 342-361, 371-390, 397-415, and 492-509 of SEQ ID NO:55, as described in Shi et al). Within transmembrane domain 4, residues 376, 379, 382, 385, and 388 are positively charged and represent the voltage sensor. A pore loop domain is also depicted (residues 463-482 of SEQ ID NO:48 and 464-482 of SEQ ID NO:55), within which residues 478-480 represent the potassium channel pore motif as described herein. A cyclic nucleotide binding domain is also depicted (residues 615-687 of SEQ ID NO:48 and SEQ ID NO:55), and, as described in Shi et al, the SDPG at residues 937-940 of SEQ ID NO:48 and residues 943-946 of SEQ ID NO:55 is a conserved motif found at the C terminal end of both mammalian *erg* protein family members.

[1640] Based on the above described sequence similarities, the 52908 molecules of the present invention belong to the potassium channel family, as described herein. Therefore, the 52908 polypeptides of the invention exhibit, and can modulate, 52908-mediated activities (e.g., potassium channel mediated activities), and can act as, or can be used to develop, novel diagnostic targets and therapeutic agents for prognosticating,

diagnosing, preventing, inhibiting, alleviating, or curing 52908-mediated or related disorders (e.g., disorders associated with potassium channel family members, as described below.

[1641] As used herein, a “potassium channel mediated activity,” includes an activity which involves a potassium channel, *e.g.*, setting the resting membrane potential, modulating the electrical activity of cells, or modulating action potential waveforms, firing frequency, and neurotransmitter and/or hormone secretion. Potassium channel mediated activities include release of neurotransmitters, *e.g.*, dopamine or norepinephrine, from cells, *e.g.*, neuronal cells; modulation of resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation; participation in signal transduction pathways, and modulation of processes such as integration of sub-threshold synaptic responses and the conductance of back-propagating action potentials in, for example, neuronal cells or muscle cells.

[1642] As used herein, a “52908 activity,” “biological activity of 52908,” or “functional activity of 52908,” refers to an activity exerted by a 52908 protein, polypeptide or nucleic acid molecule on *e.g.*, a 52908-responsive cell or on a 52908 substrate, *e.g.*, a protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 52908 activity is a direct activity, such as an association with a 52908 target molecule. A “target molecule” or “binding partner” is a molecule with which a 52908 protein binds or interacts in nature. In an exemplary embodiment, a 52908 target molecule is a 52908 ligand, *e.g.*, a potassium channel ligand (*e.g.*, a potassium channel pore-forming subunit).

[1643] A 52908 activity can also be an indirect activity, *e.g.*, a cellular signaling activity mediated by interaction of the 52908 protein with a 52908 receptor. For example, the 52908 protein of the present invention can have one or more of the following activities: (1) modulating membrane excitability; (2) modulating intracellular ion concentration; (3) modulating membrane polarization (*e.g.*, membrane polarization and/or depolarization); (4) modulating action potential; (5) modulating cellular signal transduction; (6) modulating neurotransmitter release (*e.g.*, from neuronal cells); (7) modulating synaptic transmission; (8) modulating neuronal excitability and/or plasticity; (9) modulating muscle contraction; (10) modulating cell activation (*e.g.*, T cell activation), and/or (11) modulating cellular proliferation, growth, migration and/or differentiation.

[1644] Additionally, the 52908 protein of the present invention can have one or more of the following activities: (1) interacting with a non-52908 protein molecule; (2) activating a 52908-dependent signal transduction pathway; (3) modulating membrane excitability; (4)

influencing the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation; (5) binding a cyclic nucleotide; (6) contributing to the formation of potassium channels; (7) contributing to the formation of calcium-activated, voltage independent potassium channels; (8) modulating repolarization of the neuronal cell membrane; (9) contributing to the formation of voltage-gated potassium channels; (10) contributing to the formation of cyclic nucleotide-gated potassium channels; and/or (11) modulating processes which underlie learning and memory, such as integration of sub-threshold synaptic responses and the conductance of back-propagating action potentials.

[1645] Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of, and oligopeptide uptake by cells of tissues in which 52908 molecules are expressed. Thus, the 52908 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders involving aberrant activities of these cells.

[1646] Accordingly, the 52908 molecules of the invention, as potassium channels, can mediate, and can act as novel diagnostic targets and therapeutic agents for controlling, one or more potassium channel-associated disorders, including CNS-related (e.g., neurological) disorders; pain and metabolic disorders; cardiovascular disorders; cellular proliferative, growth, differentiative, and/or migration disorders; immune, e.g., inflammatory, disorders; disorders associated with bone metabolism; endothelial cell disorders; liver disorders; and viral diseases.

[1647] As used herein, a “potassium channel-associated disorder” includes a disorder, disease or condition which is characterized by a misregulation of a potassium channel mediated activity. Potassium channel associated disorders can detrimentally affect conveyance of sensory impulses from the periphery to the brain and/or conductance of motor impulses from the brain to the periphery; integration of reflexes; interpretation of sensory impulses; cellular proliferation, growth, differentiation, or migration, and emotional, intellectual (e.g., learning and memory), or motor processes.

[1648] Examples of potassium channel associated disorders include CNS (e.g., neurological) disorders such as cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer’s disease, dementias related to Alzheimer’s disease (such as Pick’s disease), Parkinson’s and other Lewy diffuse body diseases, senile dementia, Huntington’s disease, Gilles de la Tourette’s syndrome, multiple sclerosis and multiple sclerosis variants, amyotrophic lateral sclerosis, progressive supranuclear palsy,

epilepsy, and Jakob-Creutzfeldt disease; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, e.g., amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, obsessive-compulsive disorder, psychoactive substance use disorders, panic disorder, as well as bipolar affective disorder, e.g., severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, e.g., migraine and obesity.

[1649] Further CNS (e.g., neurological) disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and

other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson's disease (paralysis agitans), corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

[1650] Potassium channel disorders also include pain and metabolic disorders. We hypothesize that this human K channel may be critical for hypersensitivity in different pain states since it is expressed in all small diameter nociceptive dorsal root ganglia (DRG)

neurons that express trkA, the receptor for nerve growth factor (NGF). NGF is a well known mediator of nociception.

[1651] The 52908 molecules of the present invention may be present on sensory neurons and, thus, may be involved in detecting, for example, noxious chemical, mechanical, or thermal stimuli and transducing this information into membrane depolarization events. Thus, the 52908 molecules by participating in pain signaling mechanisms, can modulate pain elicitation and act as targets for developing novel diagnostic targets and therapeutic agents to control pain. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, bulimia, cachexia, lipid disorders, and diabetes. Examples of pain disorders include, but are not limited to, pain response elicited during various forms of tissue injury, e.g., inflammation, infection, and ischemia, usually referred to as hyperalgesia (described in, for example, Fields, H.L., (1987) *Pain*, New York:McGraw-Hill); pain associated with musculoskeletal disorders, e.g., joint pain; tooth pain; headaches; pain associated with surgery; pain related to irritable bowel syndrome; and chest pain.

[1652] Further examples of potassium channel associated disorders include cardiac-related disorders. Cardiovascular system disorders in which the 52908 molecules of the invention may be directly or indirectly involved include arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, Jervell syndrome, Lange-Nielsen syndrome, QT syndrome (e.g., long-QT syndrome (e.g., autosomal dominant LQT-syndrome, or Romano-Ward syndrome)), congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, and arrhythmia. 52908-mediated or related disorders also include disorders of the musculoskeletal system such as paralysis and muscle weakness, e.g., ataxia, myotonia, and myokymia.

[1653] Cardiovascular disorders also include, but are not limited to cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina (e.g., angina pectoris), chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart

disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

[1654] Potassium channel disorders also include cellular proliferation, growth, differentiation, or migration disorders. Cellular proliferation, growth, differentiation, or migration disorders include those disorders that affect cell proliferation, growth, differentiation, or migration processes. As used herein, a “cellular proliferation, growth, differentiation, or migration process” is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. The 52908 molecules of the present invention can be involved in signal transduction mechanisms, which are known to be involved in cellular growth, differentiation, and migration processes. Thus, the 52908 molecules can modulate cellular growth, differentiation, or migration, and may play a role in disorders characterized by aberrantly regulated growth, differentiation, or migration.

[1655] Examples of cellular proliferation, growth, differentiation, or migration disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic

neoplastic disorders, e.g., leukemias; tumor angiogenesis and metastasis; skeletal dysplasia; neuronal deficiencies resulting from impaired neural induction and patterning; neurodegenerative disorders, e.g., Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, Jakob-Creutzfeldt disease, or AIDS related dementia; hepatic disorders; cardiovascular disorders; and hematopoietic and/or myeloproliferative disorders.

[1656] As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, e.g., malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, e.g., cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[1657] The 52908 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic

neoplastic disorders. As used herein, the term “hematopoietic neoplastic disorders” includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, *e.g.*, arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Typically, the diseases arise from poorly differentiated acute leukemias, *e.g.*, erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, L., (1991) *Crit. Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom’s macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGL), Hodgkin’s disease and Reed-Sternberg disease.

[1658] The 52908 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, *e.g.*, inflammatory (*e.g.* respiratory inflammatory), disorders. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren’s Syndrome, inflammatory bowel disease, *e.g.* Crohn’s disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener’s granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves’ disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[1659] The presence of 52908 RNA or protein can be used to identify a cell or tissue, or other biological sample, as being derived from the brain, e.g., cerebral cortex, from the heart, from a muscle, or of neuronal origin. Expression can be determined by evaluating RNA, e.g., by hybridization of a 52908 specific probe, or with a 52908 specific antibody.

[1660] The 52908 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:48 thereof are collectively referred to as “polypeptides or proteins of the invention” or “52908 polypeptides or proteins.” Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as “nucleic acids of the invention” or “52908 nucleic acids.”

[1661] As used herein, the term “nucleic acid molecule” includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[1662] The term “isolated or purified nucleic acid molecule” includes nucleic acid molecules that are separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term “isolated” includes nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an “isolated” nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5'- and/or 3'-ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kilobases, 4 kilobases, 3 kilobases, 2 kilobases, 1 kilobase, 0.5 kilobase or 0.1 kilobase of 5'- and/or 3'-nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an “isolated” nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[1663] As used herein, the term “hybridizes under stringent conditions” describes conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in available references (e.g., Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1-6.3.6). Aqueous and non-aqueous methods are described in that reference and either can be used. A preferred example of stringent

hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 50°C.

Another example of stringent hybridization conditions are hybridization in 6× SSC at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 55°C. A further example of stringent hybridization conditions are hybridization in 6× SSC at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 60°C. Preferably, stringent hybridization conditions are hybridization in 6× SSC at about 45°C, followed by one or more washes in 0.2× SSC, 0.1% (w/v) SDS at 65°C. Particularly preferred stringency conditions (and the conditions that should be used if the practitioner is uncertain about what conditions should be applied to determine if a molecule is within a hybridization limitation of the invention) are 0.5 molar sodium phosphate, 7% (w/v) SDS at 65°C, followed by one or more washes at 0.2× SSC, 1% (w/v) SDS at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:47, 54, and 57, or SEQ ID NO:49, 56, and 59, corresponds to a naturally-occurring nucleic acid molecule.

[1664] As used herein, a “naturally-occurring” nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[1665] As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules which include an open reading frame encoding 52908 protein, preferably mammalian 52908 protein, and can further include non-coding regulatory sequences and introns.

[1666] An “isolated” or “purified” polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language “substantially free” means preparation of 52908 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-52908 protein (also referred to herein as a “contaminating proteins”), or of chemical precursors or non-52908 chemicals. When the 52908 protein or biologically active portions thereof are recombinantly produced, they are also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume

of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[1667] A “non-essential” amino acid residue is a residue that can be altered from the wild-type sequence of 52908 (e.g., the sequence of SEQ ID NO:47, SEQ ID NO:49) without abolishing or, more preferably, without substantially altering a biological activity, whereas an “essential” amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the conserved potassium channel domain are predicted to be particularly non-amenable to alteration, except that amino acid residues in transmembrane domains can generally be replaced by other residues having approximately equivalent hydrophobicity without significantly altering 52908 activity.

[1668] A “conservative amino acid substitution” is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a nonessential amino acid residue in 52908 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of 52908 coding sequences, such as by saturation mutagenesis, and the resultant mutants can be screened for 52908 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:47, SEQ ID NO:49, the encoded proteins can be expressed recombinantly and the activity of the protein can be determined.

[1669] As used herein, a “biologically active portion” of 52908 protein includes fragment of 52908 protein that participate in an interaction between 52908 molecules and non-52908 molecules. Biologically active portions of 52908 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the 52908 protein, e.g., the amino acid sequences shown in SEQ ID NO:48, which include fewer amino acids than the full length 52908 protein, and exhibit at least one activity of 52908 protein. Typically, biologically active portions comprise a domain or

motif with at least one activity of the 52908 protein, e.g., the ability to modulate membrane excitability, intracellular ion concentration, membrane polarization, and action potential.

[1670] A biologically active portion of 52908 protein can be a polypeptide that is, for example, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, or more amino acids in length. Biologically active portions of 52908 protein can be used as targets for developing agents that modulate 52908-mediated activities, e.g., biological activities described herein.

[1671] Calculations of sequence homology or identity (the terms are used interchangeably herein) between sequences are performed as follows.

[1672] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 60%, 70%, 80%, 82%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the length of the reference sequence (e.g., when aligning a second sequence to the 52908 amino acid sequences of SEQ ID NO:48 having 958 amino acid residues, at least 473, preferably at least 480, more preferably at least 500, even more preferably at least 550, and even more preferably at least 560, 580, 600, 620, 640, 660, 680, 700, 720, 740, 760, 780, 800, 820, 840, 860, 880, 900, 920, 940, or 958 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[1673] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the

Needleman et al. (1970) *J. Mol. Biol.* 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package, using either a BLOSUM 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package, using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a BLOSUM 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[1674] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers et al. (1989) *CABIOS* 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[1675] The nucleic acid and protein sequences described herein can be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 52908 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 52908 protein molecules of the invention. To obtain gapped alignments for comparison purposes, gapped BLAST can be utilized as described in Altschul et al. (1997, *Nucl. Acids Res.* 25:3389-3402). When using BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[1676] 52908 polypeptides of the present invention can have amino acid sequences sufficiently or substantially identical to the amino acid sequences of SEQ ID NO:48. The terms “sufficiently identical” or “substantially identical” are used herein to refer to a first amino acid or nucleotide sequence that contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide

sequences have a common structural domain or common functional activity. For example, amino acid or nucleotide sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity are defined herein as sufficiently or substantially identical.

[1677] “Misexpression or aberrant expression”, as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over- or under-expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[1678] “Subject,” as used herein, can refer to a mammal, e.g., a human, or to an experimental animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

[1679] A “purified preparation of cells,” as used herein, refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10%, and more preferably, 50% of the subject cells.

[1680] Various aspects of the invention are described in further detail below.

ISOLATED NUCLEIC ACID MOLECULES

[1681] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes 52908 polypeptides described herein, e.g., full length 52908 protein or fragments thereof, e.g., biologically active portions of 52908 protein. Also included are nucleic acid fragments suitable for use as hybridization probes, which can be used, e.g., to identify a nucleic acid molecules encoding polypeptide of the inventions, 52908 mRNA, and

fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

[1682] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:47, or portions or fragments thereof. In one embodiment, the nucleic acid molecules include sequences encoding the human 52908 protein (i.e., “the coding region,” from nucleotides 1-2877 of SEQ ID NO:47, excluding the termination codon, shown as in SEQ ID NO:49), as well as untranslated (e.g., noncoding) sequences, e.g., 3’ untranslated sequence (i.e., 2842-3164 of SEQ ID NO:47). Alternatively, the nucleic acid molecules can include only the coding regions of SEQ ID NO:47 (e.g., 1-2877 of SEQ ID NO:49) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecules encode sequences corresponding to the mature proteins of SEQ ID NO:48. In yet another embodiment, the nucleic acid molecules encode sequences corresponding to fragments of the proteins from about amino acid 463-482 of SEQ ID NO:48.

[1683] In another embodiment, an isolated nucleic acid molecule of the invention includes nucleic acid molecules which are complements of the nucleotide sequences shown in SEQ ID NO:47, SEQ ID NO:49, or portions or fragments thereof. In other embodiments, the nucleic acid molecules of the invention are sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:47 such that they can hybridize to the nucleotide sequences shown in SEQ ID NO:47, thereby forming stable duplexes.

[1684] In one embodiment, isolated nucleic acid molecules of the present invention include nucleotide sequences which are at least about: 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more, homologous to the entire length of the nucleotide sequences shown in SEQ ID NO:47, or portions or fragments thereof, preferably of the same length, of any of these nucleotide sequences.

52908 NUCLEIC ACID FRAGMENTS

[1685] A nucleic acid molecule of the invention can include only a portion or fragment of the nucleic acid sequences of SEQ ID NO:47 or SEQ ID NO:49. For example, such a nucleic acid molecule can include fragments which can be used as probes or primers or fragments encoding a portion of 52908 protein, e.g., immunogenic or biologically active portions of 52908 protein. A fragment can comprise those nucleotides of SEQ ID NO:47

which encode a conserved domain of human 52908, e.g., a pore loop domain. The nucleotide sequences determined from the cloning of the 52908 genes allow for the generation of probes and primers designed for use in identifying and/or cloning other 52908 family members, or fragments thereof, as well as 52908 homologues, or fragments thereof, from other species.

[1686] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding or untranslated region. Other embodiments include a fragment which includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 75 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[1687] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, 52908 nucleic acid fragments can include sequences corresponding to a domain of human 52908, e.g., a pore loop domain.

[1688] 52908 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:47, SEQ ID NO:49, or of a naturally occurring allelic variant or mutant of SEQ ID NO:47 and SEQ ID NO:49.

[1689] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[1690] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes a domain of human 52908, e.g., a pore loop domain (e.g., at about nucleotides 1387-1446 of SEQ ID NO:47), or a fragment thereof.

[1691] In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 52908 sequence, e.g., a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: transmembrane domains at about amino acid residues 261-282, 295-317, 345-361, 371-390, 398-420, and 491-515 of SEQ ID NO:48; a conserved PAS domain at about amino acid residues 41-59 of SEQ ID NO:48; a conserved PAC domain at about amino acid residues 93-135 of SEQ ID NO:48; a conserved pore loop domain at about amino acid residues 463-482 of SEQ ID NO:48; a conserved cyclic nucleotide gated channel transmembrane region at about amino acid residues 341-580 of SEQ ID NO:48; and a conserved cyclic nucleotide binding domain (CNBD) at about amino acid residues 615-687 of SEQ ID NO:48.

[1692] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[1693] A nucleic acid fragment encoding a “biologically active portion of 52908 polypeptides” can be prepared by isolating a portion of the nucleotide sequences of SEQ ID NO:47 or SEQ ID NO:49, which encode polypeptides having a 52908 biological activity (e.g., the biological activities of the 52908 protein are described herein), expressing the encoded portion of the 52908 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portions of the 52908 protein. For example, nucleic acid fragments encoding biologically active portions of 52908 include a conserved domain of human 52908, e.g., a pore loop domain, e.g., amino acid residues 463-482 of SEQ ID NO:48. A nucleic acid fragment encoding a biologically active portion of a 52908 polypeptide, may comprise a nucleotide sequence which is greater than 80 or more nucleotides in length.

[1694] In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 400, 440, 480, 500, 520, 530, 560, 600, 640, 680, 700, 720, 740, 760, 780, 800, 820, 840, 860, 880, 900, 920, 940, 960, 980, 1000, 1020, 1040, 1060, 1080, 1100, 1120, 1140,

1160, 1180, 1200, 1220, 1240, 1260, 1280, 1300, 1340, 1360, 1380, 1400, 1420, 1440, 1460, 1480, 1500, 1520, 1540, 1560, 1580, 1600, 1620, 1640, 1660, 1680, 1700, 1720, 1740, 1760, 1780, 1800, 1820, 1840, 1860, 1880, 1900, 1920, 1940, 1960, 1980, 2000, 2020, 2040, 2060, 2080, 2100, 2120, 2140, 2160, 2180, 2200, 2220, 2240, 2260, 2280, 2300, 2320, 2340, 2360, 2380, 2400, 2420, 2440, 2460, 2480, 2500, 2520, 2540, 2560, 2580, 2600, 2620, 2640, 2660, 2680, 2700, 2720, 2740, 2760, 2780, 2800, 2820, 2840, 2860, 2880, 2900, 2920, 2940, 2960, 2980, 3000, 3020, 3040, 3060, 3070, or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:47, or SEQ ID NO:49, or a complement thereof.

52908 NUCLEIC ACID VARIANTS

[1695] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO:47 and SEQ ID NO:49. Such differences can be due to degeneracy of the genetic code and result in a nucleic acid which encodes the same 52908 protein as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:48. If alignment is needed for this comparison, the sequences should be aligned for maximum homology. “Looped” out sequences from deletions or insertions, or mismatches, are considered differences.

[1696] Nucleic acids of the invention can be chosen for having codons which are preferred or non-preferred for a particular expression system. For example, the nucleic acid can be one in which at least one codon, preferably at least 10% or 20% of the codons, has been altered such that the sequence is optimized for expression in bacterial (e.g., *E. coli*), yeast, human, insect, or nonhuman mammalian (e.g., CHO) cells.

[1697] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

[1698] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO:47 or and SEQ ID NO:49, e.g., as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one, but less than 1%, 5%, 10% or 20%, of the nucleotides in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. “Looped” out sequences from deletions or insertions, or mismatches, are considered differences.

[1699] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more, identical to the nucleotide sequences shown in SEQ ID NO:48, or fragments of these sequences. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to the nucleotide sequences shown in SEQ ID NO:48, or fragments of the sequences. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 52908 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 52908 genes.

[1700] Preferred variants include those that are correlated with at least one of the following 52908 biological activities: (1) modulating membrane excitability, (2) modulating intracellular ion concentration, (3) modulating membrane polarization (*e.g.*, membrane polarization and/or depolarization), (4) modulating action potential, (5) modulating cellular signal transduction, (6) modulating neurotransmitter release (*e.g.*, from neuronal cells), (7) modulating synaptic transmission, (8) modulating neuronal excitability and/or plasticity, (9) modulating muscle contraction, (10) modulating cell activation (*e.g.*, T cell activation), and/or (11) modulating cellular proliferation, growth, migration and/or differentiation.

[1701] Still other preferred variants include those that are correlated with at least one of the following 52908 biological activities: (1) interacting with a non-52908 protein molecule; (2) activating a 52908-dependent signal transduction pathway; (3) modulating membrane excitability; (4) influencing the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation; (5) binding a cyclic nucleotide; (6) contributing to the formation of potassium channels; (7) contributing to the formation of calcium-activated, voltage independent potassium channels; (8) modulating repolarization of the neuronal cell membrane; (9) contributing to the formation of voltage-gated potassium channels; (10) contributing to the formation of cyclic nucleotide-gated potassium channels;

and/or (11) modulating processes which underlie learning and memory, such as integration of sub-threshold synaptic responses and the conductance of back-propagating action potentials.

[1702] Allelic variants of 52908, e.g., human 52908, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 52908 protein within a population that maintain the 52908 biological activities described herein.

[1703] Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:48, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 52908, e.g., human 52908, protein within a population that do not demonstrate the 52908 activities described herein.

[1704] Non-functional allelic variants can typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequences of SEQ ID NO:48, or a substitutions, insertions, or deletions in critical residues or critical regions of these proteins.

[1705] Moreover, nucleic acid molecules encoding other 52908 family members and, thus, which have nucleotide sequences which differ from the 52908 sequences of SEQ ID NO:47 and SEQ ID NO:49 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 52908 Nucleic Acid Molecules

[1706] In another aspect, the invention features isolated nucleic acid molecules which are antisense to 52908. An “antisense” nucleic acid can include a nucleotide sequence which is complementary to a “sense” nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acids can be complementary to entire 52908 coding strands, or to only portions thereof (e.g., the coding regions of human 52908 corresponding to SEQ ID NO:49). In another embodiment, the antisense nucleic acid molecule is antisense to a “noncoding region” of the coding strands of nucleotide sequences encoding 52908 (e.g., the 5’ and 3’ untranslated regions).

[1707] An antisense nucleic acid can be designed such that it is complementary to the entire coding regions of 52908 mRNA, but more preferably is an oligonucleotide which is antisense to only portions of the coding or noncoding regions of 52908 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start sites of 52908 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[1708] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[1709] The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding 52908 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[1710] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2''-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

[1711] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 52908-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequences of 52908 cDNAs disclosed herein (i.e., SEQ ID NO:47 and SEQ ID NO:49), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 52908-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, 52908 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

[1712] 52908 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 52908 (e.g., the 52908 promoters and/or enhancers) to form triple helical structures that prevent transcription of the 52908 genes in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6:569-84; Helene, C. i (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[1713] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[1714] 52908 nucleic acid molecules can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. et al. (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms “peptide nucleic acid” or “PNA” refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. et al. (1996) *supra*; Perry-O’Keefe et al. *Proc. Natl. Acad. Sci.* 93: 14670-675.

[1715] PNAs of 52908 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 52908 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as “artificial restriction enzymes” when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. et al. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. et al. (1996) *supra*; Perry-O’Keefe *supra*).

[1716] In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol et al. (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[1717] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 52908 nucleic acids of the invention, two complementary regions one having a fluorophore and one a quencher

such that the molecular beacon is useful for quantitating the presence of the 52908 nucleic acids of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Patent No. 5,854,033; Nazarenko et al., U.S. Patent No. 5,866,336, and Livak et al., U.S. Patent 5,876,930.

ISOLATED 52908 POLYPEPTIDES

[1718] In another aspect, the invention features, isolated 52908 protein, or fragments, e.g., biologically active portions, for use as immunogens or antigens to raise or test (or more generally to bind) anti-52908 antibodies. 52908 protein can be isolated from cells or tissue sources using standard protein purification techniques. 52908 protein, or fragments thereof, can be produced by recombinant DNA techniques or synthesized chemically.

[1719] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when expressed the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present when expressed in a native cell.

[1720] In a preferred embodiment, 52908 polypeptides have one or more of the following characteristics:

[1721] the ability to: (1) interact with a non-52908 protein molecule; (2) activate a 52908-dependent signal transduction pathway; (3) modulate the release of neurotransmitters; (4) modulate membrane excitability; (5) influence the resting potential of membranes, wave forms and frequencies of action potentials, and thresholds of excitation; (6) bind a cyclic nucleotide; (7) contribute to the formation of potassium channels; (8) contribute to the formation of calcium-activated, voltage independent potassium channels; (9) modulate repolarization of the neuronal cell membrane; (10) contribute to the formation of voltage-gated potassium channels; (11) contribute to the formation of cyclic nucleotide-gated potassium channels; and (12) modulate processes which underlie learning and memory, such as integration of sub-threshold synaptic responses and the conductance of back-propagating action potentials.

[1722] the ability to: (1) modulate membrane excitability, (2) modulate intracellular ion concentration, (3) modulate membrane polarization (*e.g.*, membrane polarization and/or depolarization), (4) modulate action potential, (5) modulate cellular signal transduction, (6) modulate neurotransmitter release (*e.g.*, from neuronal cells), (7) modulate synaptic transmission, (8) modulate neuronal excitability and/or plasticity, (9) modulate muscle contraction, (10) modulate cell activation (*e.g.*, T cell activation), and/or (11) modulate cellular proliferation, growth, migration and/or differentiation.

[1723] a molecular weight, *e.g.*, a deduced molecular weight (*e.g.*, of 105.3 kDa), preferably ignoring any contribution of post translational modifications, amino acid composition or other physical characteristic of SEQ ID NO:48;

[1724] an overall sequence similarity of at least 50%, preferably at least 60%, more preferably at least 70, 80, 90, or 95%, with a polypeptide of SEQ ID NO:48; and

[1725] a conserved 52908 domain which is preferably about 70%, 80%, 82%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical with the sequence containing amino acid residues about 41-59, 93-135, 463-482, 341-580, and 615-687 of SEQ ID NO:48.

[1726] In a preferred embodiment, the 52908 protein, or fragments thereof, differ from the corresponding sequences in SEQ ID NO:48. In one embodiment they differ by at least one, but by less than 15, 10 or 5, amino acid residues. In another they differ from the corresponding sequences in SEQ ID NO:48 by at least one residue, but less than 20%, 15%, 10% or 5%, of the residues in them differ from the corresponding sequences in SEQ ID NO:48. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a nonessential residue or a conservative substitution. In a preferred embodiment the differences are not in the conserved potassium channel domain. In another preferred embodiment one or more differences are in the conserved potassium channel domain.

[1727] Other embodiments include a protein that contains one or more changes in amino acid sequence, *e.g.*, a change in an amino acid residue which is not essential for activity. Such 52908 protein differ in amino acid sequence from SEQ ID NO:48, yet retain biological activity.

[1728] In one embodiment, the protein includes an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 60%, 70%, 80%, 82%, 84%, 85%, 86%, 87%,

88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or more, homologous to SEQ ID NO:48.

[1729] 52908 protein or fragment is provided which varies from the sequences of SEQ ID NO:48 in regions that do not correspond to a domain specifically defined herein (e.g., from about amino acids 1 to 40 or 65 to 85) by at least one, but by less than 15, 10 or 5, amino acid residues in the protein or fragment, but which does not differ from the sequences of SEQ ID NO:48 in regions that correspond to a domain specifically defined herein (e.g., from amino acid residues about 41-59, 93-135, 463-482, 341-580, and 615-687 of SEQ ID NO:48). (If this comparison requires alignment the sequences should be aligned for maximum homology. “Looped” out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

[1730] In one embodiment, a biologically active portion of 52908 protein includes a conserved potassium channel domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of native 52908 protein.

[1731] In a preferred embodiment, the 52908 protein has amino acid sequences shown in SEQ ID NO:48. In other embodiments, the 52908 protein are substantially identical to SEQ ID NO:48. In yet another embodiment, the 52908 protein are substantially identical to SEQ ID NO:48 and retain the functional activities of the proteins of SEQ ID NO:48, as described herein.

52908 CHIMERIC OR FUSION PROTEINS

[1732] In another aspect, the invention provides 52908 chimeric or fusion proteins. As used herein, a 52908 “chimeric protein” or “fusion protein” includes a 52908 polypeptide linked to a non-52908 polypeptide. A “non-52908 polypeptide” refers to polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 52908 protein, e.g., a protein which is different from the 52908 protein and which is derived from the same or a different organism. The 52908 polypeptides of the fusion protein can correspond to all or a portion, e.g., a fragment, described herein of a 52908 amino acid sequence. In a preferred embodiment, 52908 fusion proteins include at

least one (or two) biologically active portion of 52908 protein. The non-52908 polypeptide can be fused to the N-terminus or C-terminus of the 52908 polypeptide.

[1733] One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused with the carboxyl terminus of GST sequences. Such fusion proteins can facilitate purification of a recombinant polypeptide of the invention.

[1734] In another embodiment, the fusion protein contains a heterologous signal sequence at its amino terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., supra) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

[1735] Fusion proteins can include all or a part of a serum protein, e.g., a portion of an immunoglobulin protein (e.g., IgG, IgA, or IgE); an Fc region; and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

[1736] Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-52908 antibodies directed against a polypeptide of the invention in a subject, to purify 52908 ligands and in screening assays to identify molecules which inhibit the interaction of 52908 receptors with 52908 ligands. The immunoglobulin fusion protein can, for example, comprise a portion of a polypeptide of the invention fused with the amino-terminus or the carboxyl-terminus of an immunoglobulin constant region, as disclosed in U.S. Patent No. 5,714,147, U.S. Patent No. 5,116,964, U.S. Patent No. 5,514,582, and U.S. Patent No. 5,455,165.

[1737] The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction in vivo. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand / receptor interaction can be useful therapeutically, both for treating disorders caused by, for example: (i) aberrant modification or mutation of a gene

encoding 52908 protein; (ii) mis-regulation of the 52908 genes; and (iii) aberrant post-translational modification of 52908 protein.

[1738] Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be performed using anchor primers which give rise to complementary overhangs between two consecutive gene fragments and which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see, e.g., Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

VARIANTS OF 52908 PROTEIN

[1739] In another aspect, the invention also features a variants of 52908 polypeptides, e.g., which function as agonists (mimetics) or as antagonists. Variants of the 52908 protein can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of 52908 protein. An agonist of the 52908 protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of 52908 protein. An antagonist of 52908 protein can inhibit one or more of the activities of the naturally occurring form of the 52908 protein by, for example, competitively modulating a 52908-mediated activity of 52908 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 52908 protein.

[1740] Variants of a protein of the invention which function as either agonists (e.g., mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the 52908 protein for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein

sequences can be expressed as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

[1741] In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, re-naturing the DNA to form double stranded DNA which can include sense / antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes amino terminal and internal fragments of various sizes of the protein of interest.

[1742] Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[1743] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 52908 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6:327-331).

[1744] Cell based assays can be exploited to analyze variegated 52908 libraries. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 52908 in a substrate-dependent manner. The transfected cells are then contacted with 52908 and the effect of the expression of the mutant on signaling by the 52908 substrates can be detected, for example, by assaying (i) the interaction of 52908 protein with 52908 target molecules; (ii) the interaction of 52908 protein with 52908 target molecules, wherein the 52908 target is a ligand, e.g., a potassium channel ligand; or (iii) the

interaction of 52908 protein with 52908 target molecules, wherein the 52908 target is a receptor, e.g., a potassium channel receptor. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 52908 substrate, and the individual clones further characterized.

[1745] In another aspect, the invention features a method of making 52908 polypeptides, e.g., peptides having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 52908 polypeptides, e.g., naturally occurring 52908 polypeptides. The method includes: altering the sequence of 52908 polypeptides, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[1746] In another aspect, the invention features a method of making a fragment or analog of 52908 polypeptides which demonstrate biological activities of naturally occurring 52908 polypeptides. The method includes: altering the sequence, e.g., by substitution or deletion of one or more residues, of 52908 polypeptides, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

ANTI-52908 ANTIBODIES

[1747] In another aspect, the invention provides an anti-52908 antibody. The term “antibody” as used herein refers to an immunoglobulin molecule and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as 52908 molecules. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, Fab and F(ab’)₂ fragments which can be generated by treating the antibody with an enzyme such as papain or pepsin, respectively.

[1748] The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric, humanized, fully human, non-human (e.g., murine, rat, rabbit, or goat), or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

[1749] The term “monoclonal antibody” or “monoclonal antibody composition”, as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of 52908. A

monoclonal antibody composition thus typically displays a single binding affinity for a particular 52908 protein with which it immunoreacts.

[1750] Polyclonal anti-52908 antibodies can be prepared as described above by immunizing a suitable subject with a 52908 immunogen. The anti-52908 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized 52908. If desired, the antibody molecules directed against 52908 can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-52908 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem.* 255:4980-83; Yeh et al. (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a 52908 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds 52908.

[1751] Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-52908 monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by

fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind 52908, *e.g.*, using a standard ELISA assay.

[1752] Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-52908 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with 52908 to thereby isolate immunoglobulin library members that bind 52908. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT International Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-

1377; Hoogenboom et al. (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

[1753] Additionally, chimeric, humanized, and completely human antibodies are also within the scope of the invention. Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment of human patients, and some diagnostic applications.

[1754] Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

[1755] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93; and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

[1756] Completely human antibodies that recognize a selected epitope can be generated using a technique referred to as “guided selection.” In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described by Jespers et al. (1994) *Bio/Technology* 12:899-903).

[1757] Full-length 52908 protein, or antigenic peptide fragments of 52908, can be used as an immunogen or can be used to identify anti-52908 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptides of 52908 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:48 and encompass an epitope of 52908. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[1758] Fragments of 52908 which include, e.g., residues 463-482 of SEQ ID NO:48, can be used as immunogens to make an antibodies against the conserved 52908 domains of the invention, e.g., potassium channel domains (e.g., a pore loop domain).

[1759] Antibodies reactive with, or specific or selective for, any of these regions, or other regions or domains described herein are provided.

[1760] In an alternative embodiment, the antibody fails to bind to an Fc receptor, e.g., it is a type which does not support Fc receptor binding or has been modified, e.g., by deletion or other mutation, such that it does not have a functional Fc receptor binding region.

[1761] Preferred epitopes encompassed by the antigenic peptide are regions of 52908 which are located on the surface of the protein, e.g., hydrophilic regions (depicted, e.g., in a hydropathy plot), as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 52908 protein sequences can be used to identify the regions that have a particularly high probability of being localized to the surface of the 52908 protein, and are thus likely to constitute surface residues useful for targeting antibody production.

[1762] In a preferred embodiment the antibody binds an epitope on any domain or region on 52908 protein described herein.

[1763] The anti-52908 antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered as described, for example, in Colcher, D. et al., (1999)

Ann. NY Acad. Sci. 880: 263-80; and Reiter, Y., *Clin. Cancer Res.* 1996 Feb;2(2):245-52.

The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 52908 protein.

[1764] Anti-52908 antibodies (e.g., monoclonal antibodies) can be used to isolate 52908 proteins by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-52908 antibody can be used to detect 52908 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-52908 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[1765] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

[1766] A vector can include 52908 nucleic acids in a form suitable for expression of the nucleic acids in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be

expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 52908 protein, mutant forms of 52908 protein, fusion proteins, and the like).

[1767] The recombinant expression vectors of the invention can be designed for expression of 52908 protein in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA . Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[1768] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech, Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly MA) and pRIT5 (Pharmacia, Piscataway NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[1769] Purified fusion proteins can be used in 52908 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 52908 protein. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

[1770] To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[1771] The 52908 expression vectors can be yeast expression vectors, vectors for expression in insect cells, e.g., baculovirus expression vectors or vectors suitable for expression in mammalian cells.

[1772] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[1773] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166).

Developmentally-regulated promoters are also encompassed, for example, the murine hox

promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[1774] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., (1986) Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics* 1:1.

[1775] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 52908 nucleic acid molecule within a recombinant expression vector or a 52908 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[1776] A host cell can be any prokaryotic or eukaryotic cell. For example, 52908 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[1777] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[1778] A host cell of the invention can be used to produce (i.e., express) 52908 protein. Accordingly, the invention further provides methods for producing 52908 protein using the host cells of the invention. In one embodiment, the method includes culturing the

host cell of the invention (into which a recombinant expression vector encoding 52908 protein has been introduced) in a suitable medium such that 52908 protein is produced. In another embodiment, the method further includes isolating 52908 protein from the medium or the host cell.

[1779] In another aspect, the invention features a cell or a purified preparation of cells which includes 52908 transgenes, or which otherwise misexpresses 52908. The cell preparation can consist of human or nonhuman cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell, or cells, include 52908 transgenes, e.g., a heterologous form of 52908, e.g., a gene derived from humans (in the case of a non-human cell). The 52908 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell, or cells, includes a gene which misexpresses an endogenous 52908, e.g., a gene, the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed 52908 alleles or for use in drug screening.

[1780] In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes subject 52908 polypeptides.

[1781] Also provided are cells, preferably human cells, e.g., human hematopoietic or fibroblast cells, in which endogenous 52908 genes are under the control of a regulatory sequence that does not normally control the expression of the endogenous 52908 genes. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 52908 genes. For example, an endogenous 52908 genes which are "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991.

Transgenic Animals

[1782] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of 52908 protein and for identifying and/or evaluating modulators of 52908 activity. As used herein, a "transgenic animal" is a non-

human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which endogenous 52908 genes have been altered by, e.g., by homologous recombination between the endogenous genes and exogenous DNA molecules introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

[1783] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of 52908 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 52908 transgene in its genome and/or expression of 52908 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding 52908 protein can further be bred to other transgenic animals carrying other transgenes.

[1784] 52908 protein or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[1785] The invention also includes a population of cells from a transgenic animal, as discussed, e.g., herein.

Uses

[1786] The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays;

b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

[1787] The isolated nucleic acid molecules of the invention can be used, for example, to express 52908 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect 52908 mRNA (e.g., in a biological sample) or a genetic alteration in 52908 genes, and to modulate 52908 activity, as described further below. The 52908 protein can be used to treat disorders characterized by insufficient or excessive production of a 52908 substrate or production of 52908 inhibitors. In addition, the 52908 protein can be used to screen for naturally occurring 52908 substrates, to screen for drugs or compounds which modulate 52908 activity, as well as to treat disorders characterized by insufficient or excessive production of 52908 protein or production of 52908 protein forms which have decreased, aberrant or unwanted activity compared to 52908 wild type protein. Moreover, the anti-52908 antibodies of the invention can be used to detect and isolate 52908 protein, regulate the bioavailability of 52908 protein, and modulate 52908 activity.

[1788] A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 52908 polypeptide is provided. The method includes: contacting the compound with the subject 52908 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 52908 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 52908 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 52908 polypeptide. Screening methods are discussed in more detail herein.

Screening Assays:

[1789] The invention provides methods (also referred to herein as “screening assays”) for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 52908 protein, have a stimulatory or inhibitory effect on, for example, 52908 expression or 52908 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 52908 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 52908 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[1790] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of 52908 protein or polypeptides or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of 52908 protein or polypeptides or a biologically active portion thereof.

[1791] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann, R.N. et al. (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the “one-bead one-compound” library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

[1792] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop et al. (1994) *J. Med. Chem.* 37:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, USP 5,223,409), spores (Ladner USP “409), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra.*).

[1793] In one embodiment, an assay is a cell-based assay in which a cell which expresses 52908 protein or biologically active portions thereof are contacted with a test compound, and the ability of the test compound to modulate 52908 activity is determined. Determining the ability of the test compound to modulate 52908 activity can be accomplished by monitoring, for example, (i) the interaction of 52908 protein with a 52908

target molecule; (ii) the interaction of 52908 protein with a 52908 target molecule, wherein the 52908 target is a potassium channel substrate. The cell, for example, can be of mammalian origin, e.g., human.

[1794] The ability of the test compound to modulate 52908 binding to a compound, e.g., a 52908 substrate, or to bind to 52908 can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to 52908 can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, 52908 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate 52908 binding to a 52908 substrate in a complex. For example, compounds (e.g., 52908 substrates) can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[1795] The ability of a compound (e.g., a 52908 substrate) to interact with 52908, with or without the labeling of any of the interactants, can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with 52908 without the labeling of either the compound or the 52908. McConnell, H. M. et al. (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and 52908.

[1796] In yet another embodiment, a cell-free assay is provided in which 52908 protein, or biologically active portion thereof, is contacted with a test compound and the ability of the test compound to bind to the 52908 protein, or biologically active portion thereof, is evaluated. Preferred biologically active portions of the 52908 protein to be used in assays of the present invention include fragments which participate in interactions with non-52908 molecules, e.g., fragments with high surface probability scores.

[1797] Soluble and/or membrane-bound forms of isolated proteins (e.g., 52908 protein, or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to

utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton[®] X-100, Triton[®] X-114, Thesit[®], Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

[1798] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[1799] The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz et al., U.S. Patent No. 5,631,169; Stavrianopoulos, et al., U.S. Patent No. 4,868,103). A fluorophore label on the first, “donor” molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, “acceptor” molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the “donor” protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the “acceptor” molecule label may be differentiated from that of the “donor”. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the “acceptor” molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[1800] In another embodiment, determining the ability of the 52908 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705). “surface plasmon resonance” or “BIA” detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[1801] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[1802] It may be desirable to immobilize 52908, an anti-52908 antibody, or a 52908 target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to 52908 protein, or interaction of 52908 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/52908 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 52908 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 52908 binding or activity determined using standard techniques.

[1803] Other techniques for immobilizing either 52908 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 52908 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[1804] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where

the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[1805] In one embodiment, this assay is performed utilizing antibodies reactive with 52908 protein or target molecules but which do not interfere with binding of the 52908 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 52908 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 52908 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 52908 protein or target molecule.

[1806] Alternatively, cell-free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A.P., (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. et al., eds. *Current Protocols in Molecular Biology* 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. et al., eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N.H., (1998) *J Mol Recognit* 11:141-8; Hage, D.S., and Tweed, S.A. (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[1807] In a preferred embodiment, the assay includes contacting the 52908 protein, or biologically active portion thereof, with a known compound which binds 52908 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with 52908 protein, wherein determining the ability of the test compound to interact with 52908 protein includes determining the ability of the

test compound to preferentially bind to 52908, or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

[1808] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as “binding partners.” Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to, molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 52908 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of 52908 protein through modulation of the activity of a downstream effector of a 52908 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[1809] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[1810] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding

partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described herein.

[1811] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[1812] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[1813] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending

upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[1814] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[1815] In yet another aspect, the 52908 protein can be used as a “bait protein” in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 52908 (“52908-binding proteins” or “52908-bp”) and are involved in 52908 activity. Such 52908-bps can be activators or inhibitors of signals by the 52908 protein or 52908 targets as, for example, downstream elements of a 52908-mediated signaling pathway.

[1816] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for 52908 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein (“prey” or “sample”) is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: 52908 protein can be the fused to the activator domain.) If the “bait” and the “prey” proteins are able to interact, *in vivo*, forming a 52908-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and

used to obtain the cloned gene which encodes the protein which interacts with the 52908 protein.

[1817] In another embodiment, modulators of 52908 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 52908 mRNA or protein evaluated relative to the level of expression of 52908 mRNA or protein in the absence of the candidate compound. When expression of 52908 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 52908 mRNA or protein expression. Alternatively, when expression of 52908 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 52908 mRNA or protein expression. The level of 52908 mRNA or protein expression can be determined by methods described herein for detecting 52908 mRNA or protein.

[1818] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of 52908 protein can be confirmed *in vivo* in an animal model.

[1819] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 52908 modulating agent, an anti-sense 52908 nucleic acid molecule, a 52908-specific antibody, or a 52908-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

[1820] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 52908 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[1821] The 52908 nucleotide sequences or portions thereof can be used to map the location of the 52908 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 52908 sequences with genes associated with disease.

[1822] Briefly, 52908 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 52908 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 52908 sequences will yield an amplified fragment.

[1823] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al. (1983) *Science* 220:919-924).

[1824] Other mapping strategies e.g., *in situ* hybridization (described in Fan, Y. et al. (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 52908 to a chromosomal location.

[1825] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., *Human Chromosomes: A Manual of Basic Techniques* ((1988) Pergamon Press, New York).

[1826] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[1827] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al. (1987) *Nature*, 325:783-787.

[1828] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 52908 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[1829] 52908 sequences can be used to identify individuals from biological samples using, e.g., restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, e.g., in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[1830] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 52908 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[1831] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:47 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those in SEQ ID NO:49 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[1832] If a panel of reagents from 52908 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

Use of Partial 52908 Sequences in Forensic Biology

[1833] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[1834] The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:47 (e.g., fragments derived from the noncoding regions of SEQ ID NO:47 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[1835] The 52908 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 52908 probes can be used to identify tissue by species and/or by organ type.

[1836] In a similar fashion, these reagents, e.g., 52908 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

[1837] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[1838] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 52908.

[1839] Such disorders include, e.g., a disorder associated with the misexpression of 52908 genes.

[1840] The method includes one or more of the following: (i) detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 52908 genes, or detecting the presence or absence of a mutation in a region which controls the expression of the genes, e.g., a mutation in the 5' control region; (ii) detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 52908 genes; (iii) detecting, in a tissue of the subject, the misexpression of the 52908 genes, at the mRNA level, e.g., detecting a non-wild type level of a mRNA; or (iv) detecting, in a tissue of the subject, the misexpression of the genes, at the protein level, e.g., detecting a non-wild type level of a 52908 polypeptides.

[1841] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 52908 genes; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

[1842] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which

hybridizes to a sense or antisense sequence from SEQ ID NO:47, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 52908 genes; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[1843] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 52908 genes; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 52908.

[1844] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[1845] In preferred embodiments the method includes determining the structure of 52908 genes, an abnormal structure being indicative of risk for the disorder.

[1846] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the 52908 protein or nucleic acids, which hybridizes specifically with the genes. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

[1847] The presence, level, or absence of 52908 protein or nucleic acids in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 52908 protein or nucleic acids (e.g., mRNA, genomic DNA) that encodes 52908 protein such that the presence of 52908 protein or nucleic acids are detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 52908 genes can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 52908 genes; measuring the amount of protein encoded by the 52908 genes; or measuring the activity of the protein encoded by the 52908 genes.

[1848] The level of mRNA corresponding to 52908 genes in a cell can be determined both by *in situ* and by *in vitro* formats.

[1849] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction

analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, full-length 52908 nucleic acids, such as the nucleic acid of SEQ ID NO:47, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 52908 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

[1850] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 52908 genes.

[1851] The level of mRNA in a sample that is encoded by one of 52908 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli et al., (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh et al., (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al., (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi et al., U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[1852] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 52908 gene being analyzed.

[1853] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 52908 mRNA, or genomic DNA, and comparing the presence of 52908 mRNA or genomic DNA in the control sample with the presence of 52908 mRNA or genomic DNA in the test sample.

[1854] A variety of methods can be used to determine the level of protein encoded by 52908. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[1855] The detection methods can be used to detect 52908 protein in a biological sample *in vitro*, as well as *in vivo*. *In vitro* techniques for detection of 52908 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 52908 protein include introducing into a subject a labeled anti-52908 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[1856] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 52908 protein, and comparing the presence of 52908 protein in the control sample with the presence of 52908 protein in the test sample.

[1857] The invention also includes kits for detecting the presence of 52908 in a biological sample. For example, the kit can include a compound or agent capable of detecting 52908 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 52908 protein or nucleic acids.

[1858] For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the

invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[1859] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention, or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[1860] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 52908 expression or activity. As used interchangeably herein, the terms “unwanted” and “undesirable” include an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

[1861] In one embodiment, a disease or disorder associated with aberrant or unwanted 52908 expression or activity is identified. A test sample is obtained from a subject and 52908 protein or nucleic acids (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 52908 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 52908 expression or activity. As used herein, a “test sample” refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

[1862] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 52908 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a cellular proliferative and/or differentiative disorder, a hormonal disorder, an immune or

inflammatory disorder, a neurological disorder, a cardiovascular disorder, a blood vessel disorder, or a platelet disorder.

[1863] The methods of the invention can also be used to detect genetic alterations in 52908 genes, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 52908 protein activity or nucleic acid expression, such as a cellular proliferative and/or differentiative disorder, a hormonal disorder, an immune or inflammatory disorder, a neurological disorder, a cardiovascular disorder, a blood vessel disorder, or a platelet disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding 52908 protein, or the mis-expression of the 52908 genes. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from 52908 genes; 2) an addition of one or more nucleotides to 52908 genes; 3) a substitution of one or more nucleotides of 52908 genes, 4) a chromosomal rearrangement of 52908 genes; 5) an alteration in the level of a messenger RNA transcript of 52908 genes, 6) aberrant modification of 52908 genes, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of 52908 genes, 8) a non-wild type level of 52908 protein, 9) allelic loss of 52908 genes, and 10) inappropriate post-translational modification of 52908 protein.

[1864] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 52908 gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to 52908 genes under conditions such that hybridization and amplification of the 52908 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

[1865] In another embodiment, mutations in 52908 genes from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example,

sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[1866] In other embodiments, genetic mutations in 52908 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotides probes (Cronin, M.T. et al. (1996) *Human Mutation* 7: 244-255; Kozal, M.J. et al. (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 52908 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. et al. *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[1867] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 52908 gene and detect mutations by comparing the sequence of the sample 52908 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry.

[1868] Other methods for detecting mutations in the 52908 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242; Cotton et al. (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295).

[1869] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 52908 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[1870] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 52908 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 52908 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

[1871] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[1872] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki et al. (1986) *Nature* 324:163; Saiki et al. (1989) *Proc. Natl Acad. Sci USA* 86:6230).

[1873] Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[1874] The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving 52908 genes.

USE OF 52908 MOLECULES AS SURROGATE MARKERS

[1875] The 52908 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 52908 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the 52908 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the

presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen et al. (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[1876] The 52908 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a “pharmacodynamic marker” is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug in vivo. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g., a 52908 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-52908 antibodies may be employed in an immune-based detection system for 52908 protein marker, or 52908-specific radiolabeled probes may be used to detect 52908 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda et al. US 6,033,862; Hattis et al.

(1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[1877] The 52908 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a “pharmacogenomic marker” is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod et al. (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 52908 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 52908 DNA may correlate 52908 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

[1878] The nucleic acid and polypeptides, fragments thereof, as well as anti-52908 antibodies and small molecule modulators of 52908 molecules (also referred to herein as “active compounds”) of the invention can be incorporated into pharmaceutical compositions. Such compositions typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, a “pharmaceutically acceptable carrier” includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[1879] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile

diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[1880] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin. Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum

drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[1881] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[1882] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[1883] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[1884] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[1885] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be

obtained commercially from Alza Corporation (Palo Alto CA) and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[1886] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[1887] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1888] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[1889] As defined herein, a therapeutically effective amount of protein or polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body

weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[1890] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[1891] The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[1892] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per

kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[1893] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive agent (e.g., a radioactive metal ion). A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[1894] The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue

plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[1895] Techniques for conjugating a therapeutic moiety to an antibody are well known (see, e.g., Arnon et al., 1985, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in *Monoclonal Antibodies And Cancer Therapy*, Reisfeld et al., Eds., Alan R. Liss, Inc. pp. 243-256; Hellstrom et al., 1987, "Antibodies For Drug Delivery", in *Controlled Drug Delivery*, 2nd ed., Robinson et al., Eds., Marcel Dekker, Inc., pp. 623-653; Thorpe, 1985, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in *Monoclonal Antibodies "84: Biological And Clinical Applications*, Pinchera et al., Eds., pp. 475-506; "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in *Monoclonal Antibodies For Cancer Detection And Therapy*, Baldwin et al., Eds., Academic Press, pp. 303-316, 1985; and Thorpe et al., 1982, *Immunol. Rev.*, 62:119-158). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[1896] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[1897] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment

[1898] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or undesirable 52908 expression or activity. With regard to both prophylactic

and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics.

“Pharmacogenomics”, as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and commercially available. More specifically, the term refers the study of how a patient’s genes determine his or her response to a drug (e.g., a patient’s “drug response phenotype”, or “drug response genotype”.) Thus, another aspect of the invention provides methods for tailoring an individual’s prophylactic or therapeutic treatment with either the 52908 molecules of the present invention or 52908 modulators according to that individual’s drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to identify patients who will experience toxic drug-related side effects.

[1899] “Treatment,” as used herein, is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, palliate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[1900] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or undesirable 52908 expression or activity, by administering to the subject a 52908 molecule or an agent which modulates 52908 expression or at least one 52908 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or undesirable 52908 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 52908 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 52908 aberrance, for example, a 52908 molecule (e.g., a 52908 nucleic acid molecule or 52908 protein or polypeptide, or a fragment thereof, as described herein), or 52908 agonist or 52908 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[1901] It is possible that some 52908 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[1902] As discussed, successful treatment of 52908 disorders can be brought about by techniques that serve to inhibit the expression or activity of 52908 target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 52908 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, human, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab'')₂ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[1903] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[1904] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[1905] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by 52908 expression is through the use of aptamer molecules specific for 52908 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to protein ligands (see, e.g., Osborne, et al. (1997) *Curr. Opin. Chem. Biol.* 1(1):5-9; and Patel, D.J. (1997) *Curr. Opin. Chem. Biol.*

1(1):32-46). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic protein molecules may be, aptamers offer a method by which 52908 protein activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[1906] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 52908 disorders. For a description of antibodies, see the Antibody section above.

[1907] In circumstances wherein injection of an animal or a human subject with 52908 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 52908 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D. (1999) *Ann. Med.* 31(1):66-78; and Bhattacharya-Chatterjee, M., and Foon, K.A. (1998) *Cancer Treat. Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 52908 protein. Vaccines directed to a disease characterized by 52908 expression may also be generated in this fashion.

[1908] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[1909] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 52908 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

[1910] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for

determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[1911] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[1912] Another measurement which can be used to determine the effective dose for an individual is to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 52908 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique is found in Ansell, R. J. et al. (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea, K.J. (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrices in this way can be seen in Vlatakis, G. et al., (1993) *Nature* 361:645-647. Through the use of isotope-labeling,

the “free” concentration of compound which modulates the expression or activity of 52908 can be readily monitored and used in calculations of IC_{50} .

[1913] Such “imprinted” affinity matrices can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50} . A rudimentary example of such a “biosensor” is discussed in Kriz, D. et al. (1995) *Analytical Chemistry* 67:2142-2144.

[1914] Another aspect of the invention pertains to methods of modulating 52908 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 52908 molecule (e.g., a 52908 nucleic acid molecule or 52908 protein or polypeptide, or a fragment thereof, as described herein) or an agent that modulates one or more of the activities of the 52908 protein activity associated with the cell. An agent that modulates 52908 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of 52908 protein (e.g., a 52908 substrate, ligand, or receptor), an anti-52908 antibody, a 52908 agonist or antagonist, a peptidomimetic of a 52908 agonist or antagonist, or other small molecule.

[1915] In one embodiment, the agent stimulates one or more 52908 activities. Examples of such stimulatory agents include active 52908 protein and nucleic acid molecules encoding a 52908 protein or polypeptide, or a fragment thereof. In another embodiment, the agent inhibits one or more 52908 activities. Examples of such inhibitory agents include antisense 52908 nucleic acid molecules, anti-52908 antibodies, and 52908 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject), or *in situ*. As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 52908 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) 52908 expression or activity. In another embodiment, the method involves administering a 52908 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or undesirable 52908 expression or activity.

[1916] Stimulation of 52908 expression or activity is desirable in situations in which 52908 expression or activity is abnormally downregulated and/or in which increased 52908 expression or activity is likely to have a beneficial effect. Likewise, inhibition of 52908 expression or activity is desirable in situations in which 52908 expression or activity is abnormally upregulated and/or in which decreased 52908 expression or activity is likely to have a beneficial effect.

[1917] The 52908 molecules can act as novel diagnostic targets and therapeutic agents for controlling one or more of CNS-related (e.g., neurological) disorders; pain and metabolic disorders; cardiovascular disorders; cellular proliferative, growth, differentiative, and/or migration disorders; immune, e.g., inflammatory, disorders; disorders associated with bone metabolism; endothelial cell disorders; liver disorders; and viral diseases.

[1918] Aberrant expression and/or activity of 52908 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 52908 molecules in bone cells, e.g. osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, 52908 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 52908 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[1919] As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, e.g., proliferation, migration, angiogenesis, or vascularization; or aberrant expression of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic

retinopathy, endometriosis, Grave's disease, ischemic disease (*e.g.*, atherosclerosis), and chronic inflammatory diseases (*e.g.*, rheumatoid arthritis).

[1920] Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of preexisting fibers. The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (*e.g.*, bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (*e.g.*, storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (*e.g.*, tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (*e.g.*, Zellweger syndrome). Additionally, the methods described herein can be useful for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyldopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[1921] Additionally, 52908 molecules can play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 52908 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 52908 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

Pharmacogenomics

[1922] The 52908 molecules of the present invention, as well as agents, and modulators which have a stimulatory or inhibitory effect on a 52908 activity (e.g., 52908 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 52908 associated disorders (e.g., cellular proliferative and/or differentiative disorders, hormonal disorders, immune and inflammatory disorders, neurological disorders, cardiovascular disorders, blood vessel disorders, and platelet disorders) associated with aberrant or undesirable 52908 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 52908 molecule or 52908 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 52908 molecule or 52908 modulator.

[1923] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. et al. (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985 and Linder, M.W. et al. (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[1924] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be

compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a “SNP” is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

[1925] Alternatively, a method termed the “candidate gene approach”, can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug’s target is known (e.g., 52908 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[1926] Alternatively, a method termed the “gene expression profiling”, can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 52908 molecule or 52908 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[1927] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 52908 molecule or 52908 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[1928] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 52908 genes of the present invention,

wherein these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 52908 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells, will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

[1929] Monitoring the influence of agents (e.g., drugs) on the expression or activity of 52908 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 52908 gene expression or protein levels, or upregulate 52908 activity, can be monitored in clinical trials of subjects exhibiting decreased 52908 gene expression or protein levels, or downregulated 52908 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 52908 gene expression or protein levels, or downregulate 52908 activity, can be monitored in clinical trials of subjects exhibiting increased 52908 gene expression or protein levels, or upregulated 52908 activity. In such clinical trials, the expression or activity of 52908 genes, and preferably, other genes that have been implicated in, for example, a 52908-associated disorder can be used as a “read out” or markers of the phenotype of a particular cell.

OTHER EMBODIMENTS

[1930] In another aspect, the invention features a method of analyzing a plurality of capture probes. The method is useful, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence, wherein the capture probes are from a cell or subject which expresses 52908 or from a cell or subject in which a 52908 mediated response has been elicited; contacting the array with a 52908 nucleic acid (preferably purified), a 52908 polypeptide (preferably purified), or an anti-52908 antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by a signal generated from a label attached to the 52908 nucleic acid, polypeptide, or antibody.

[1931] The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

[1932] The method can include contacting the 52908 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

[1933] The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 52908. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

[1934] The method can be used to detect SNPs, as described above.

[1935] In another aspect, the invention features, a method of analyzing 52908, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 52908 nucleic acid or amino acid sequence; comparing the 52908 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 52908.

[1936] The method can include evaluating the sequence identity between a 52908 sequence and a database sequence. The method can be performed by accessing the database at a second site, e.g., over the internet. Preferred databases include GenBank™ and SwissProt.

[1937] In another aspect, the invention features, a set of oligonucleotides, useful, e.g., for identifying SNP's, or identifying specific alleles of 52908. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, e.g., an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotides which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second allele.

[1938] The sequence of a 52908 molecules is provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 52908 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, e.g., an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

[1939] A 52908 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having thereon 52908 sequence information of the present invention.

[1940] As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus of other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

[1941] As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 52908 sequence information.

[1942] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a 52908 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-

available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[1943] By providing the 52908 nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[1944] The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder, wherein the method comprises the steps of determining 52908 sequence information associated with the subject and based on the 52908 sequence information, determining whether the subject has a 52908-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1945] The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a 52908-associated disease or disorder or a pre-disposition to a disease associated with 52908, wherein the method comprises the steps of determining 52908 sequence information associated with the subject, and based on the 52908 sequence information, determining whether the subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

[1946] The present invention also provides in a network, a method for determining whether a subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder, said method comprising the steps of receiving 52908 sequence information from the subject and/or information related thereto, receiving

phenotypic information associated with the subject, acquiring information from the network corresponding to 52908 and/or corresponding to a 52908-associated disease or disorder, and based on one or more of the phenotypic information, the 52908 information (e.g., sequence information and/or information related thereto), and the acquired information, determining whether the subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1947] The present invention also provides a business method for determining whether a subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder, said method comprising the steps of receiving information related to 52908 (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 52908 and/or related to a 52908-associated disease or disorder, and based on one or more of the phenotypic information, the 52908 information, and the acquired information, determining whether the subject has a 52908-associated disease or disorder or a pre-disposition to a 52908-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[1948] The invention also includes an array comprising a 52908 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 52908. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

[1949] In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is

useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[1950] In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a 52908-associated disease or disorder, progression of 52908-associated disease or disorder, and processes, such a cellular transformation associated with the 52908-associated disease or disorder.

[1951] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of 52908 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[1952] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including 52908) that could serve as a molecular target for diagnosis or therapeutic intervention.

[1953] As used herein, a “target sequence” can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[1954] Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used

in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

[1955] Thus, the invention features a method of making a computer readable record of a sequence of a 52908 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[1956] In another aspect, the invention features, a method of analyzing a sequence. The method includes: providing a 52908 sequence, or record, in computer readable form; comparing a second sequence to the 52908 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, e.g., determining if the 52908 sequence includes a sequence being compared. In a preferred embodiment the 52908 or second sequence is stored on a first computer, e.g., at a first site and the comparison is performed, read, or recorded on a second computer, e.g., at a second site. E.g., the 52908 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

[1957] This invention is further illustrated by the following examples, which should not be construed as limiting.

EXAMPLES

GENE EXPRESSION ANALYSIS (EXPERIMENT I)

[1958] Total RNA was prepared from various human tissues by a single step extraction method using RNA STAT-60 according to the manufacturer's instructions (TelTest, Inc). Each RNA preparation was treated with DNase I (Ambion) at 37°C for 1 hour. DNase I treatment was determined to be complete if the sample required at least 38 PCR amplification cycles to reach a threshold level of fluorescence using β -2 microglobulin

as an internal amplicon reference. The integrity of the RNA samples following DNase I treatment was confirmed by agarose gel electrophoresis and ethidium bromide staining. After phenol extraction cDNA was prepared from the sample using the SUPERSCRIPT™ Choice System following the manufacturer's instructions (GibcoBRL). A negative control of RNA without reverse transcriptase was mock reverse transcribed for each RNA sample.

[1959] Human 52908 expression was measured by TaqMan® quantitative PCR (Perkin Elmer Applied Biosystems) in cDNA prepared from several human and rat tissue and cell line panels, containing at least the following normal or diseased tissues or cell lines: normal artery; diseased aorta; normal vein; coronary smooth muscle cells; HUVEC (human umbilical vein endothelial cells); hemangioma; normal heart; congestive heart failure heart; normal kidney; skeletal muscle; normal adipose tissue; pancreas; primary osteoblasts; osteoclasts; normal skin; normal spinal cord; normal brain cortex; normal brain hypothalamus; nerve; dorsal root ganglia; normal breast; normal ovary; ovary tumor; normal prostate; prostate tumor; salivary glands; normal colon; colon tumor; normal lung; lung COPD; normal liver; liver fibrosis; normal spleen; normal tonsil; normal lymph node; normal small intestine; macrophages; synovium; BM-MNC; activated peripheral blood mononuclear cells; neutrophils; megakaryocytes; erythroid; and a positive control.

[1960] Probes were designed by PrimerExpress software (PE Biosystems) based on the sequence of the human 52908 gene.

[1961] Each human 52908 gene probe was labeled using FAM (6-carboxyfluorescein), and the β 2-microglobulin reference probe was labeled with a different fluorescent dye, VIC. The differential labeling of the target gene and internal reference gene thus enabled measurement in same well. Forward and reverse primers and the probes for both β 2-microglobulin and target gene were added to the TaqMan® Universal PCR Master Mix (PE Applied Biosystems). Although the final concentration of primer and probe could vary, each was internally consistent within a given experiment. A typical experiment contained 200nM of forward and reverse primers plus 100nM probe for β -2 microglobulin and 600 nM forward and reverse primers plus 200 nM probe for the target gene. TaqMan matrix experiments were carried out on an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems).

[1962] The following method was used to quantitatively calculate human 52908 gene expression in the various tissues relative to β -2 microglobulin expression in the same tissue. The threshold cycle (Ct) value is defined as the cycle at which a statistically

significant increase in fluorescence is detected. A lower Ct value is indicative of a higher mRNA concentration. The Ct value of the human 52908 gene is normalized by subtracting the Ct value of the β -2 microglobulin gene to obtain a Δ Ct value using the following formula: Δ Ct = Ct_{human 52908} - Ct _{β -2 microglobulin}. Expression is then calibrated against a cDNA sample showing a comparatively low level of expression of the human 52908 gene. The Δ Ct value for the calibrator sample is then subtracted from Δ Ct for each tissue sample according to the following formula: $\Delta\Delta$ Ct = Δ Ct_{sample} - Δ Ct_{calibrator}. Relative expression is then calculated using the arithmetic formula given by $2^{-\Delta\Delta$ Ct}. Expression of the target human 52908 gene in each of the tissues tested is then graphically represented.

[1963] Human 52908 mRNA expression in various tissues and cell lines as described above, relative to a positive control, is as follows:

[1964] In the human panel, expression is highest in dorsal root ganglia, followed by prostate and brain.

[1965] In one rat panel, the highest level of expression is in dorsal root ganglia, followed by much lower expression in spinal cord and brain.

[1966] In another rat panel, 52908 is not regulated in dorsal root ganglia after capsaicin treatment, but is upregulated in the dorsal horn of the spinal cord..

GENE EXPRESSION ANALYSIS- *In Situ* (EXPERIMENT II)

[1967] ISH hybridization with a human and rat probes shows very low levels of expression of this gene in rat brain. In the spinal cord, this gene is expressed in sub-population of neurons in laminae I, II and V. Also, some expression is observed around the central canal, lamina X. Confirming the Taqman data, very high levels of expression of 52908 were detected in a sub-population of nociceptive dorsal root ganglia neurons. This neuronal population corresponds to that of small and intermediate diameter.

EQUIVALENTS

[1968] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.